

Grau en Estadística

Títol: Anàlisi de l'IBEX 35 i predicció mitjançant tècniques de *Machine Learning*

Autor: Raquel Jiménez Iranzo

Director: Francisco Javier Sierra Martínez

Departament: Econometria, Estadística i Economia Aplicada

Convocatòria: 2n Semestre del curs 2023 (Juny 2023)



AGRAÏMENTS

Vull agrair primer de tot a la meva família, per la seva paciència i suport des de l'inici dels estudis fins al final.

Als meus amics i amigues, per ajudar-me a evadir-me quan era necessari. En especial al Martí, per estar en els moments més difícils al meu costat.

Sense oblidar-me del Javier, el qual m'ha animat des del principi i ha dedicat el seu temps en compartir amb mi les seves idees per tal d'ajudar-me i treure una bona feina.

Gràcies!

RESUM

L'ús de tècniques de *Machine Learning*, cada dia és més prevalent en diferents àmbits i sectors, incloent-hi l'anàlisi i la predicció dels mercats financers com l'IBEX 35. Aquest treball, té com a objectiu principal analitzar les dades històriques de l'índex IBEX 35 obtingudes de *Yahoo Finance*, des de l'any 2000 fins al 2022, per poder conèixer el comportament que ha tingut al llarg del temps. Aquestes dades seran preparades pel seu posterior entrenament utilitzant tres eines diferents de *Machine Learning*, les quals són les xarxes neuronals artificials, el *random forest* i el *gradient boosting*, per poder predir l'IBEX 35. Un cop implementats els models, s'avaluaran i es seleccionarà el mètode que té millor capacitat predictiva, comentant també els avantatges i inconvenients de cadascun.

PARAULES CLAU

IBEX 35, *Machine Learning*, Predicció, Xarxes Neuronals Artificials, *Random Forest*, *Gradient Boosting*.

RESUMEN

El uso de técnicas de Machine Learning, cada día es más prevalente en diferentes ámbitos y sectores, incluyendo el análisis y la predicción de los mercados financieros como el IBEX 35. Este trabajo, tiene como objetivo principal analizar los datos históricos del índice IBEX 35 obtenidas de *Yahoo Finance*, desde el año 2000 hasta el 2022, para poder conocer el comportamiento que ha tenido a lo largo del tiempo. Estos datos serán preparados para su posterior entrenamiento utilizando tres herramientas diferentes de *Machine Learning*, las cuales son redes neuronales artificiales, *random forest* y *gradient boosting*, para poder predecir el IBEX 35. Una vez implementados los modelos, se evaluarán y se seleccionará el método que tiene mejor capacidad predictiva, comentando también las ventajas e inconvenientes de cada uno.

PALABRAS CLAVE

IBEX 35, *Machine Learning*, Predicción, Redes Neuronales Artificiales, *Random Forest*, *Gradient Boosting*.

TITLE: IBEX 35 analysis and prediction using Machine Learning techniques

ABSTRACT

The use of Machine Learning techniques is becoming more prevalent every day in different fields and sectors, including the analysis and prediction of financial markets such as the IBEX 35. The main objective of this work is to analyze the historical data of the IBEX 35 index obtained from Yahoo Finance from the year 2000 to 2022, in order to know the behavior that it has had over time. These data will be prepared for subsequent training using three different Machine Learning tools, which are artificial neural networks, random forest and gradient boosting, in order to predict the IBEX 35. Once the models have been implemented, they will be evaluated to determine which method has better predictive capacity, as well as comment on the advantages and disadvantages of each one.

KEYWORDS

IBEX 35, Machine Learning, Prediction, Artificial Neural Networks, Random Forest, Gradient Boosting.

CLASSIFICACIO AMS

62M10 Time series, auto-correlation, regression, etc

62M20 Prediction

62M45 Neural nets and related approaches

62P05 Applications to actuarial sciences and financial mathematics

91B84 Economic time series analysis

TAULA DE CONTINGUTS

| | |
|---|-----------|
| 1. INTRODUCCIÓ | 10 |
| 1.1. <i>Justificació</i> | <i>10</i> |
| 1.2. <i>Objectius.....</i> | <i>11</i> |
| 1.3. <i>Estructura del treball.....</i> | <i>11</i> |
| 2. MARC TEÒRIC..... | 13 |
| 2.1. <i>Mercats financers.....</i> | <i>13</i> |
| 2.2. <i>La borsa de valors.....</i> | <i>16</i> |
| 2.3. <i>Introducció a la borsa espanyola i l'IBEX 35.....</i> | <i>18</i> |
| 2.4. <i>Predicció de la borsa.....</i> | <i>25</i> |
| 3. METODOLOGIA..... | 29 |
| 3.1. <i>Llenguatge de programació</i> | <i>29</i> |
| 3.2. <i>Font de les dades.....</i> | <i>29</i> |
| 3.3. <i>Sèries temporals</i> | <i>31</i> |
| 3.4. <i>Machine Learning.....</i> | <i>32</i> |
| 4. MACHINE LEARNING: EINES UTILITZADES..... | 37 |
| 4.1. <i>Xarxes neuronals artificials.....</i> | <i>37</i> |
| 4.2. <i>Random forest.....</i> | <i>40</i> |
| 4.3. <i>Gradient boosting.....</i> | <i>43</i> |
| 4.4. <i>Mètodes d'avaluació</i> | <i>44</i> |
| 4.4.1. <i>Mean Square Error (MSE) i Root Mean Square Error (RMSE)</i> | <i>44</i> |
| 4.4.2. <i>Mean Absolute Error (MAE) i Mean Absolute Percentage Error (MAPE)</i> | <i>45</i> |
| 5. ESTUDI DE LES DADES | 46 |
| 5.1. <i>Estructura de les dades</i> | <i>46</i> |
| 5.2. <i>Tractament de les dades.....</i> | <i>48</i> |
| 5.3. <i>Anàlisi descriptiva.....</i> | <i>49</i> |
| 5.4. <i>Tractament de missings.....</i> | <i>53</i> |
| 5.5. <i>Tractament d'outliers</i> | <i>55</i> |

| | | |
|-----------------|--|-----------|
| 5.6. | <i>Estadístics principals</i> | 56 |
| 5.6.1. | <i>Mitjanes mòbils</i> | 57 |
| 5.6.2. | <i>Índex de força relativa (RSI)</i> | 62 |
| 5.7. | <i>Preparació de les dades per utilitzar els mètodes de Machine Learning</i> | 63 |
| 6. | MODELATGE | 64 |
| 6.1. | <i>Xarxes neuronals artificials</i> | 64 |
| 6.2. | <i>Random forest</i> | 65 |
| 6.3. | <i>Gradient boosting</i> | 66 |
| 7. | RESULTATS | 68 |
| 7.1. | <i>Xarxes neuronals artificials</i> | 68 |
| 7.2. | <i>Random forest</i> | 70 |
| 7.3. | <i>Gradient boosting</i> | 71 |
| 8. | AVALUACIÓ DELS RESULTATS OBTINGUTS | 74 |
| 8.1. | <i>Selecció del millor mètode</i> | 75 |
| 9. | CONCLUSIONS | 76 |
| 10. | BIBLIOGRAFIA | 78 |
| ANNEX I. | Scripts d'RStudio | 82 |

ÍNDIX DE FIGURES

| | |
|--|----|
| Figura 2.1: El vell lloc de borsa d'Amsterdam (1670), per Job Adriaensz Berckheyde | 17 |
| Figura 2.2: Índexs que conformen l'IBEX..... | 20 |
| Figura 2.3: Tipus d'anàlisi borsaris..... | 25 |
| Figura 2.4: Exemple de nivells de suport i resistència..... | 26 |
| Figura 3.1: Pàgina web Yahoo Finance | 30 |
| Figura 3.2: Dades de l'IBEX 35 a Yahoo Finance | 30 |
| Figura 3.3: Exemple de sèrie temporal..... | 31 |
| Figura 3.4: Diferents tipus d'aprenentatge automàtic | 33 |
| Figura 3.5: Algoritmes d'aprenentatge automàtic supervisat | 34 |
| Figura 3.6: Algoritmes d'aprenentatge automàtic no supervisat | 35 |
| Figura 3.7: Tipus d'intel·ligència artificial | 35 |
| Figura 4.1: Comparativa xarxa neuronal humana i artificial..... | 37 |
| Figura 4.2: Estructura d'una xarxa neuronal artificial..... | 38 |
| Figura 4.3: Estructura dels arbres de decisió..... | 41 |
| Figura 4.4: Estructura del random forest..... | 41 |
| Figura 4.5: Estructura del gradient boosting | 43 |
| Figura 5.1: Primers registres de la base de dades importada | 47 |
| Figura 5.2: Sèrie temporal IBEX 35 | 48 |
| Figura 5.3. Primers registres de la base de dades un cop realitzat el tractament..... | 49 |
| Figura 5.4: Gràfic de les variables Open, Close, High, Low i Adjusted..... | 50 |
| Figura 5.5: Gràfic de la variable Volume..... | 51 |
| Figura 5.6: Gràfic de la variable Volume fins l'any 2010..... | 51 |
| Figura 5.7: Gràfic de la variable Volume entre el 2002 i el 2010..... | 52 |
| Figura 5.8: Matriu de correlació | 53 |
| Figura 5.9: Boxplot de la variable Adjusted | 56 |
| Figura 5.10: Mitjana mòbil simple de la variable Adjusted | 58 |
| Figura 5.11: Mitjana mòbil exponencial de la variable Adjusted | 59 |
| Figura 5.12: Mitjana mòbil ponderada de la variable Adjusted | 60 |
| Figura 5.13: Comparació mitjanes mòbils de la variable Adjusted | 61 |
| Figura 5.14: Gràfic de l'índex de força relativa | 62 |
| Figura 7.1: Gràfic d'importància relativa de les variables predictores | 68 |
| Figura 7.2: Xarxa neuronal resultant | 69 |
| Figura 7.3: Gràfic de la predicció de la variable Adjusted amb xarxes neuronals | 70 |
| Figura 7.4: Gràfic de la predicció de la variable Adjusted amb random forest | 70 |
| Figura 7.5: Gràfic de la influència relativa de les variables..... | 71 |
| Figura 7.6: Gràfic de la predicció de la variable Adjusted amb gradient boosting..... | 72 |
| Figura 7.7: Gràfic de la influència relativa de les variables..... | 72 |
| Figura 7.8: Gràfic de la predicció de la variable Adjusted amb gradient boosting..... | 73 |

ÍNDIX DE TAULES

| | |
|---|----|
| Taula 2.1: Cinc borses de valors més importants del món per capitalització de mercat..... | 18 |
| Taula 2.2: Empreses que conformen l'IBEX 35 | 23 |
| Taula 2.3: Recompte d'empreses que conformen l'IBEX 35 per sector..... | 24 |
| Taula 2.4: Cinc empreses amb més ponderació a l'IBEX 35..... | 24 |
| Taula 5.1: Variables de la base de dades | 46 |
| Taula 5.2: Summary de les variables | 49 |
| Taula 5.3: Dates en les quals es troben missings..... | 54 |
| Taula 5.4: Comparació mitjanes | 55 |
| Taula 5.5: Errors de la mitjana mòbil simple | 58 |
| Taula 5.6: Errors de la mitjana mòbil exponencial | 59 |
| Taula 5.7: Errors de la mitjana mòbil ponderada | 61 |
| Taula 5.8: Comparació dels errors | 61 |
| Taula 8.1: Errors de les diferents tècniques de Machine Learning utilitzades | 74 |

1. INTRODUCCIÓ

En aquest treball de final de grau s'aborda l'anàlisi de l'IBEX 35 i seva la predicció mitjançant tècniques de *Machine Learning*, també conegudes com a eines d'aprenentatge automàtic. L'IBEX 35, un dels índexs borsaris principals a Espanya, desperta un gran interès en l'àmbit financer i econòmic, ja que el seu comportament influeix en les decisions d'inversió i en l'estabilitat del mercat.

1.1. Justificació

La justificació d'aquest estudi és principalment la importància d'analitzar i predir el comportament de l'índex IBEX 35, un dels més importants en l'àmbit espanyol. L'IBEX 35 és un referent clau en el mercat financer, a més de tenir un impacte significatiu a l'economia nacional.

L'anàlisi d'aquest proporciona informació valuosa sobre les tendències i els cicles del mercat. Permet identificar patrons històrics i comprendre millor com ha evolucionat al llarg del temps. A més, és de gran utilitat per detectar senyals anticipats de canvis o tendències, que poden ajudar a predir possibles moviments futurs de l'índex.

L'ús de tècniques de *Machine Learning* per la predicció de l'IBEX 35 ofereix diversos avantatges. En primer lloc, aquestes eines permeten processar gran volum de dades històriques i extreure patrons que a simple vista no es podrien veure, per tant, ens pot proporcionar una visió més precisa i detallada.

D'altra banda, les tècniques d'aprenentatge automàtic tenen capacitat d'adaptar-se i aprendre de les dades en temps real. Això, permet generar pronòstics actualitzats i més precisos, i és de gran importància en un entorn financer caracteritzat per la volatilitat i la ràpida evolució del mercat.

No obstant això, tant l'anàlisi com la predicció de l'IBEX 35 són imprescindibles pels inversors, gestors de carteres i altres parts interessades en la presa de decisions financeres. És de gran utilitat, concretament pels inversors, ja que els permet trobar indicadors i senyals que els ajuda a ajustar les seves estratègies d'inversió per tal de maximitzar els guanys i minimitzar els riscos.

A banda dels arguments anteriors, l'elecció d'aquest tema també ve justificada per l'interès que tinc sobre les eines de *Machine Learning* i la seva utilitat. Penso que al llarg del grau d'Estadística he après molt, i aquest m'ha enriquit tant com estudiant com persona. Però, si hagués de quedar-me amb alguna assignatura amb la qual seguir els meus estudis seria Mètodes Estadístics en Minería de Dades. En aquesta, hem après mètodes de *Machine*

Learning, fet que m'ha semblat molt interessant i m'ha impulsat a voler estudiar i profunditzar en aquest tema.

A més, el preu de les accions i, de manera general, el mercat financer, sempre m'han semblat atractiu i considero que no tenia gaires coneixements al respecte.

Per tant, el treball final de grau era l'eina perfecta per ajuntar dos temes els quals m'agradaven i, d'aquesta manera, aprofitar la realització d'aquest per nodrir-me de coneixement i experiència en la recerca.

1.2. Objectius

El primer objectiu d'aquest treball consisteix en adquirir un coneixement profund sobre l'índex IBEX 35, la composició, la importància en el mercat financer i la influència en l'economia espanyola. Això implica analitzar la seva història i conèixer les principals característiques sobre aquest.

Atès que l'estudi es basa en tècniques de *Machine Learning*, el segon objectiu es basa en adquirir una entesa sòlida dels conceptes, algorismes i enfocaments utilitzats en aquesta àrea. Això inclou comprendre els diferents tipus d'algorismes d'aprenentatge automàtic, les aplicacions i les limitacions, així com les millors pràctiques per al preprocessament de dades i l'avaluació de models.

Un cop adquirit el coneixement sobre *Machine Learning*, el següent objectiu és investigar i comparar els resultats dels diferents mètodes i algorismes usats per la predicció, i veure la qualitat d'aquests.

Basant-se en la comparació anterior, l'objectiu final és identificar i seleccionar el mètode o algorisme de *Machine Learning* que demostrï ser el més adequat i eficient per a la predicció de l'IBEX 35. Això implica avaluar la precisió, la robustesa i l'estabilitat dels diferents models, així com considerar factors com ara el temps d'entrenament, la complexitat i la interpretabilitat del model seleccionat.

1.3. Estructura del treball

Entorn l'estructura que tindrà aquest treball final de grau, constarà vuit grans seccions, les quals s'expliquen a continuació.

Primerament, trobem l'apartat del marc teòric on s'examinen els fonaments dels mercats financers, s'explica el funcionament de la borsa i se'n proporciona una descripció detallada de

l'índex IBEX 35, destacant-ne la rellevància i l'impacte en l'economia. A més, es revisen les teories i tècniques existents per a la predicció de la borsa.

Tot seguit trobem la secció de la metodologia utilitzada a l'estudi, on s'especifica el llenguatge de programació utilitzat, la font de les dades utilitzades, com també l'enfocament utilitzat per a l'anàlisi de sèries temporals i la visió general dels algorismes i tècniques de *Machine Learning*.

La tercera secció es basa en l'estudi de les dades usades, on es duu a terme una anàlisi exploratòria, incloent-hi visualitzacions, estadístiques descriptives per comprendre millor la naturalesa de les dades i el tractament d'aquestes per poder realitzar posteriorment les prediccions.

En quart lloc, trobem les eines utilitzades de *Machine Learning*, on es detallen els algorismes i les tècniques seleccionades per la implementació de la predicció.

Després hi ha el modelatge emprat, on s'expliquen els models que s'han dut a terme, com s'han entrenat i ajustat els paràmetres pertinents.

La sisena secció conté els resultats obtinguts de les prediccions de l'IBEX 35, presentats de manera clara i concisa per poder tenir una millor comprensió d'aquests.

A continuació hi trobem l'avaluació dels resultats obtinguts, on s'examina la validesa i el rendiment dels models. A més, es discuteixen les limitacions de les diferents tècniques emprades i es comparen entre elles, per poder d'aquesta manera seleccionar la millor entre les utilitzades per predir l'índex.

Tot seguit, hi ha la secció de conclusions generals del treball. Es recapitularan els objectius establerts i s'avaluaran si s'han assolit, a més de resumir els resultats principals i conclusions derivades de l'estudi.

Per acabar, es trobaran els apartats de bibliografia i annex. La bibliografia contindrà les diferents fonts emprades al llarg de l'estudi. D'altra banda, a l'annex estarà tot el codi d'RStudio necessari per dur a terme l'anàlisi i les prediccions de l'IBEX 35.

2. MARC TEÒRIC

2.1. Mercats financers

Un mercat financer és un espai (virtual, físic o ambdós) el qual s'utilitza per realitzar intercanvis i negociar la compravenda d'instruments financers i definir les seves preferències. En termes generals, un mercat que negocia béns podria qualificar-se com un mercat financer, si la intenció del comprador no és consumir el producte immediatament, sinó ajornar el consum en el temps (col·laboradors de *Wikipedia*, 2022).

Els mercats financers possibiliten l'augment de capital, concretament en els mercats de capitals. En els mercats de derivats, es facilita la transferència del risc i, finalment, el comerç internacional es proporciona en el mercat de divises.

A continuació, s'observa el llistat de les principals funcions dels mercats financers:

- Identificar els mecanismes que simplifiquen la comunicació entre els participants en negociació.
- Establir els preus dels productes financers, en funció de la seva oferta i demanda.
- Reduir els costos d'intermediació, que gràcies a això facilita una major circulació dels productes.
- Gestionar els fluxos de liquiditat de productes o mercat determinat cap a altres.

Entorn les característiques que defineixen els mercats financers, és important recalcar que són diferents en funció del tipus de producte negociat. Així i tot, trobem que tenen en comú les següents:

Amplitud

El mercat financer és un espai en el qual es negocia una àmplia varietat de títols financers. L'amplitud fa referència d'aquests que es comercialitzen. En general, com més ampli sigui el mercat financer, més gran serà la diversitat d'opcions d'inversió disponibles per als participants.

Profunditat

Un altre aspecte important és la profunditat. Aquest, es refereix a l'existència dels títols financers que estiguin per sobre i per sota del preu d'equilibri que existeix en un instant i, d'aquesta manera, que cobreixin diverses eventualitats del mercat.

Llibertat

La característica de llibertat és deguda al fet que no existeix cap barrera a l'entrada o sortida del mercat financer. Al ser així, els participants poden prendre decisions d'inversió amb major llibertat i flexibilitat. D'altra banda, pot fomentar la competència i la innovació del mateix mercat.

Flexibilitat

La flexibilitat és un altre aspecte fonamental en el mercat financer, referent a la rapidesa amb què els compradors i venedors de títols poden reaccionar davant els canvis dels preus dels actius financers que es produeixen a l'economia. Així doncs, permet als participants adaptar-se a aquests canvis i prendre decisions informades d'inversió.

Transparència

Per acabar, la transparència és un concepte clau, ja que ofereix la possibilitat d'obtenir informació clara i precisa sobre el preu de l'actiu financer, permetent d'aquesta manera als participants prendre decisions d'inversió ben fonamentades. Aquest aspecte, pot fomentar la confiança en el mercat financer i garantir la integritat de les operacions realitzades.

Els mercats financers són una part fonamental de l'economia mundial i compren una àmplia gamma de productes i serveis financers. Entre els més coneguts es troben:

Mercat monetari

El conformen mercats a l'engròs, on s'intercanvien actius de gran fluïdesa. Està enfocat al curt termini (fins a un any), i el risc és baix.

Mercat de capitals

Facilita a les empreses i els governs l'obtenció de capital a mitjà i llarg termini per finançar les seves inversions. Aquest, inclou els mercats de renda variables (referent a les accions), renda fixa i bons (referent als deutes).

Mercat de derivats

Mercat ampli i complex, on es negocien instruments financers que tenen un valor dependent d'altres actius. Són els actius amb major risc de pèrdues elevades, ja que permet el palanquejament.

Mercat de divises

També conegut com a Forex, fixa el tipus de canvi de divises i on s'efectuen les operacions d'intercanvi d'aquestes en un termini menor a dos dies. Es considera el mercat més gran del món.

Aquests mercats financers són essencials per l'economia, ja que permeten a les empreses i els governs obtenir finançament, com també als inversors invertir els seus diners per obtenir beneficis. Cadascun d'aquests mercats té les seves pròpies característiques, per tant, és molt important comprendre bé el funcionament individual abans d'invertir en ells.

L'objectiu principal que es persegueix en el present treball és la predicció de l'índex IBEX 35. En conseqüència, s'enfoca en particular en el mercat de capitals. Concretament, ens fixarem en el de rendes variables.

El mercat de renda variable, també conegut com a mercat d'accions o mercat borsari, fa referència a les borses de valors. És aquell on es compren i es venen valors emesos per empreses, amb l'objectiu d'aconseguir beneficis mitjançant la variació dels preus.

Aquest mercat té diverses particularitats que el fan únic i atractiu pels inversors. Per començar, el caracteritza la volatilitat. Aquest tipus de mercat és altament volàtil i, conseqüentment, els preus de les accions poden fluctuar ràpidament i de forma significativa en un període curt de temps. Això, és degut a la gran quantitat de factors que influeixen en el preu de les accions, com per exemple, la situació econòmica, la política, les notícies corporatives, entre d'altres.

D'altra banda, el determina la rendibilitat. És un mercat que ofereix altes possibilitats de rendibilitat a llarg termini. Els inversors poden obtenir beneficis significatius si compren accions d'empreses, les quals tinguin un bon exercici financer, sempre que estiguin ben gestionades. A més, la probabilitat d'obtenir rendibilitat és més elevada en comparació amb la renda fixa.

Aquest tipus de mercat és molt líquid, és a dir, els inversors poden comprar i vendre accions de manera fàcil i ràpida. Això, permet ajustar les seves posicions en el mercat i aprofitar les oportunitats d'inversió.

La diversificació és una altra característica important, ja que permet als inversors diversificar la seva cartera d'inversions mitjançant la compra d'accions de diferents empreses i sectors econòmics. Així doncs, es redueix el risc de pèrdua i s'augmenten les possibilitats d'assolir una rendibilitat atractiva.

Finalment, cal destacar que és un mercat transparent, on es publica tota la informació rellevant sobre les empreses cotitzades i les seves activitats financeres. A més, es distingeix per la seva facilitat d'accés i la baixa necessitat de capital per poder intervenir en aquest.

Com hem comentat anteriorment, els mercats financers estan afectats per la llei de l'oferta i la demanda, cosa que suposa moviments de tipus irracional a causa dels agents humans que intervenen, com per exemple, iniciatives, pors, estratègies i condicionants. Això provoca gran una dificultat per aconseguir predir el moviment del mercat.

2.2. La borsa de valors

Una borsa de valors és un mercat físic o virtual on els demandants i oferents de capital poden posar-se en contacte, com també els que fan transaccions a través d'intermediaris autoritzats. Més popularment es coneix com la borsa (Roldán, 2017).

En el mercat borsari trobem tres participants principals: les empreses, els inversors i els intermediaris. Les empreses emeten accions per tal d'obtenir finançament i d'aquesta manera poder créixer. Aquestes, són comprades pels inversors, els quals es converteixen en propietaris d'una part del capital de l'empresa. I per acabar hi ha els intermediaris, que són cases de borsa, agents de borsa, societats de corretatge i borsa, societats de valors i agències de valors i borsa, tots amb un mateix objectiu: contactar amb els que desitgen comprar i vendre, i perfeccionar la respectiva operació rebent a canvi dels seus serveis una comissió.

A l'hora d'invertir, existeixen diferents tipus de perfils d'inversors, cadascun amb la seva pròpia estratègia i objectius. Aquests, es poden classificar en quatre categories diferents: inversors en funció de la seva aversió al risc, inversors en funció del subjecte de la inversió, inversors en funció de l'objectiu de la inversió i, inversors en funció de la manera d'operar.

Tot i existir moltes formes d'invertir, els classificaríem en els tres tipus més importants: els inversors a curt termini (traders o especuladors), els inversors a llarg termini i els inversors adversos al risc.

Els inversors a curt termini són aquells que busquen maximitzar els beneficis en un període curt de temps, generalment mitjançant la compra i venda dels actius financers en un termini de dies, setmanes o mesos. Aquests inversors són coneguts per assumir grans riscos i buscar altes rendibilitats en poc temps. L'estratègia sol ser arriscada i no és adequada per tothom degut a les grans pèrdues que pot provocar en cas que el mercat es mogui a contracorrent.

D'altra banda, els inversors a llarg termini són aquells que busquen rendibilitat a través de dividends, ampliacions de capital o altres possibles estratègies en un temps més prolongat. Aquests, busquen invertir amb una visió de futur i en empreses sòlides i estables, les quals

tinguin perspectives de creixement i generin beneficis sostenibles al llarg del temps. El tipus d'estratègia és menys arriscada que l'anterior.

Per últim, els inversors adversos al risc dirigeixen les seves inversions cap a productes de renda fixa, en molts casos, de rendibilitat baixa. Aquests busquen la seguretat i estabilitat en les seves inversions i, no es disposen a assumir grans riscos. Els productes solen ser bons o dipòsits bancaris, els quals no ofereixen una rendibilitat alta.

El terme que generalment uneix els inversors amb les empreses són les accions. Les accions són les parts iguals en què es divideix el capital social d'una societat anònima. Aquestes parts esmentades són posseïdes per una persona, l'accionista, i representen la propietat que la persona té de l'empresa, concretament, el percentatge de l'empresa que pertany a l'accionista. El preu d'aquestes ve determinat per una sèrie de factors, com l'exercici financer de l'empresa, la seva posició al mercat, l'estat de l'economia, les decisions dels reguladors governamentals i les expectatives dels mateixos inversors.

Per conèixer l'origen de la borsa de valors ens situem a la ciutat de Bruges, Bèlgica. Una família de banquers, anomenada Van der Buërse, organitzava en el seu palau reunions on es transaccionaven actius i es duïen a terme operacions comercials. Va ser en aquell moment que va aparèixer el terme borsa. Més concretament, va sorgir de l'escut familiar, el qual contenia tres bosses de pell.

No obstant això, la primera borsa de valors va sorgir l'any 1460 a la ciutat d'Anvers del mateix país. La segona, es va crear a Amsterdam, a principis del segle XVII. Aquesta, és considerada la més antiga del món, fundada el 1602 per la *Compañía holandesa de las Indias Orientales*. Va ser la primera a negociar formalment actius financers. Posteriorment, es va crear la de Londres (1570), Lió (1595) i Nova York (1792).



Figura 2.1: El vell lloc de borsa d'Amsterdam (1670), per Job Adriaensz Berckheyde (Font: Wikipedia)

Actualment, cada país compta amb la seva pròpia borsa de valors, tot i que n'hi ha que destaquen més que d'altres en termes de capitalització de mercat i volum de negociació. A continuació, es presenten les cinc principals borses de valors del món per capitalització de mercat.

Taula 2.1: Cinc borses de valors més importants del món per capitalització de mercat

| P. | País | Borsa | Capitalització |
|-----------|------------------------|---------------------------------------|-----------------------|
| 1 | Estats Units d'Amèrica | <i>New York Stock Exchange (NYSE)</i> | 25,15 bilions de \$ |
| 2 | Estats Units d'Amèrica | <i>NASDAQ Stock Exchange (NASDAQ)</i> | 17,81 bilions de \$ |
| 3 | Xina | <i>Shanghai Stock Exchange (SSE)</i> | 7,15 bilions de \$ |
| 4 | Japó | <i>Tokyo Stock Exchange (JPX)</i> | 5,4 bilions de \$ |
| 5 | Xina | <i>Shenzhen Stock Exchange (SZSE)</i> | 5,1 bilions de \$ |

(Font: Trading hours)

Podem observar com les dues borses de valors més importants del món pertanyen als Estats Units. Entre elles, la diferència és d'uns 7 bilions de dòlars. La tercera més rellevant és la de Xangai, la qual té una capitalització de 18 bilions de dòlars menys que la primera. Per acabar, les dues restants, que pertanyen a Tòquio i Shenzhen, estant en el quart i cinquè lloc respectivament, tenen una capitalització similar, al voltant de 5 bilions de dòlars.

Veiem com, tot i ser les cinc més rellevants, la borsa de Nova York té molta diferència a la resta. Cadascuna d'aquestes borses té les seves pròpies característiques i sectors clau en els quals s'enfoquen, fet que les converteix en destins destacats pels inversors de tot el món.

2.3. Introducció a la borsa espanyola i l'IBEX 35

El mercat borsari a Espanya comença el 10 de setembre de 1831, juntament amb la publicació de la Llei de Creació de la Borsa de Madrid, que va donar origen a una institució que avui dia és un referent mundial. Des de llavors, la borsa madrilenya ha evolucionat constantment, adaptant-se als canvis de l'entorn econòmic i financer.

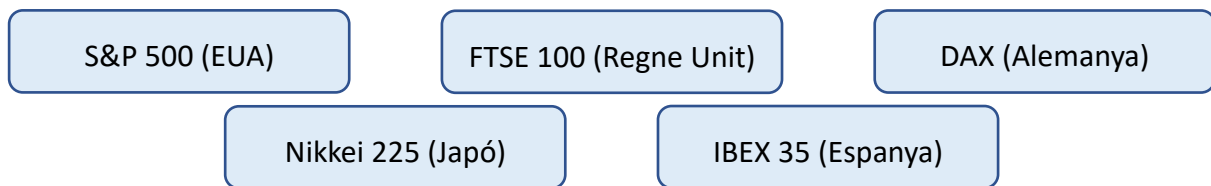
Al llarg dels anys, s'han anat creant altres borses de valors a diferents ciutats del país, com la Borsa General de Comerç de Bilbao, constituïda el 1890, i la Borsa Oficial de Comerç de

Barcelona, constituïda el 1915. Per últim, es va crear la de València, que es va unir al grup el 1980. Aquestes quatre borses van conformar el mercat borsari espanyol.

Per facilitar la interconnexió i l'intercanvi d'informació entre les diferents borses, es va crear el SIBE (Sistema d'Interconnexió Borsari Espanyol), el qual permet la negociació de valors en temps real i la cotització d'accions en totes les borses del país. D'aquesta manera, el mercat borsari espanyol s'ha convertit en un espai dinàmic i eficient per la inversió i el creixement empresarial. A més, s'ha consolidat com un dels més importants d'Europa.

Cada borsa de valors té un índex borsari, el qual és una mesura estadística per mostrar l'evolució en el temps dels preus de les accions cotitzades en aquell mercat de valors o dels seus rendiments. Aquests són molt importants, ja que donen molta informació sobre com s'està desenvolupant el mercat. També són una eina útil pels inversors, analistes i gestors de fons d'inversió, perquè els permet avaluar la rendibilitat i el risc que té, i d'aquesta manera, comparar-lo amb altres mercats. D'altra banda, es poden utilitzar per dissenyar estratègies a llarg termini, inclús per crear carteres imitant el comportament del mateix índex.

Alguns dels índexs borsaris més coneguts en l'àmbit mundial són:



L'IBEX 35 és l'índex borsari de referència de la borsa espanyola. Aquest, mesura el comportament d'un conjunt de 35 empreses que cotitzen a la Borsa de Madrid, les quals compleixen millor els paràmetres de capitalització, liquiditat i volum negociat.

Es tracta d'un índex ponderat per capitalització borsària, el que significa que les empreses de major capitalització tenen més pes a l'índex que les empreses que en tenen menor. Això és degut al fet que el càlcul de l'índex es basa en el nombre d'accions en circulació per preu d'acció. Per tant, el valor d'aquest es veu influenciat principalment pel comportament de les empreses més grans i negociades.

A més de ser un indicador de l'exercici del mercat borsari espanyol, l'IBEX 35 també és un bon indicador de l'estat de l'economia espanyola. Sí aquest puja, pot indicar que l'economia està en bona forma i que les empreses que el conformen estan en creixement. D'altra banda, si l'índex baixa, pot ser un indicati que l'economia espanyola estigui experimentant dificultats.

Aquest, va néixer el 14 de gener de 1992 per servir de referència al mercat d'opcions i futurs a Espanya. Durant aquests 30 anys d'història, l'IBEX 35 ha estat fonamental per l'economia espanyola, ja que ha estat capaç de triplicar la mida en termes de PIB, com també el valor de les empreses de l'indicador s'ha multiplicat per deu.

En aquestes tres dècades, ha consolidat la seva posició com a marca líder i principal indicador del mercat borsari espanyol, el que ha permès als inversors i analistes utilitzar-lo com a referent per avaluar les empreses cotitzades a Espanya.

Tot i haver passat més de 60 companyies diferents per l'IBEX 35, aquest índex ha estat capaç de mantenir la seva rellevància i posició de lideratge. Això és pel fet que s'actualitza regularment per reflectir els canvis en el mercat.

És interessant saber que dins de l'IBEX hi ha una família d'índexs, i no només existeix l'IBEX 35, sent el primer en néixer. A continuació es mostra els diferents tipus que hi ha:

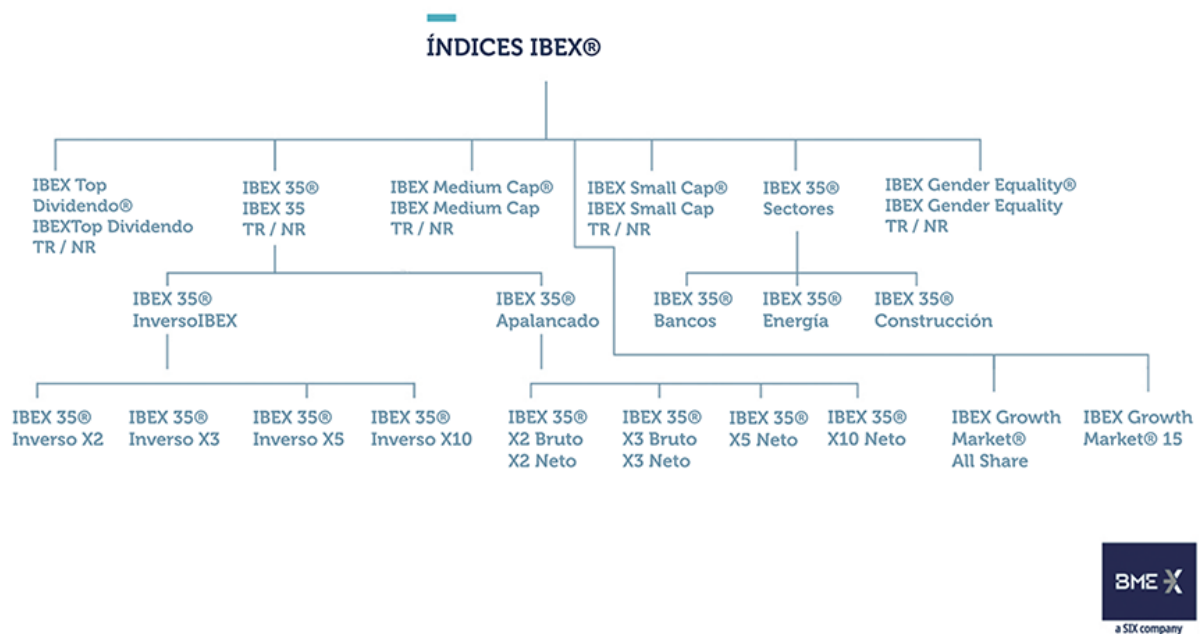


Figura 2.2: Índexs que conformen l'IBEX
(Font: Bolsas y mercados españoles - BME)

Una pregunta interessant seria la següent: com s'escullen les empreses que el conformen l'IBEX 35? Aquestes, són seleccionades per un grup d'experts anomenat *Comité Asesor Técnico* (CAT), els quals es reuneixen dos cops a l'any (cada sis mesos) i s'encarreguen de supervisar i discutir el pes de les companyies en el mercat borsari.

Per poder dur-ho a terme, es tenen en compte factors com:

- El preu unitari de cada acció.
- La variació en percentatge i en euros respecte a la sessió anterior.
- El volum d'accions que s'intercanvien en una sessió, especificat en euros.
- La capitalització, és a dir, la dimensió econòmica de l'empresa (nombre d'accions en circulació multiplicat pel preu de cadascuna).
- El PER, un indicador que indica la rendibilitat de la inversió al llarg de l'any si aquest s'hagués mantingut constant.
- La rendibilitat per dividend: el resultat de dividir el dividend esperat entre el preu de cada acció.

Per la selecció, les empreses han de tenir una sèrie de criteris objectius com per exemple la capitalització borsària, la liquiditat, el volum de negociació i la presència en el mercat. En concret, per poder accedir a formar part del conjunt d'empreses, han de complir els següents requisits:

- Ha de ser una societat anònima cotitzada a la Borsa de Madrid.
- Haver estat objecte de contractació en el mercat continu durant almenys els sis mesos anteriors a la revisió de la composició de l'índex.
- Tenir una rendibilitat mitjana superior al 0,3% de la capitalització mitjana de l'IBEX-35 durant el període de control.
- El valor ha d'haver estat negociat almenys en un terç de les sessions del període de control.
- Per últim, en cas que la negociació no funcionés, es pot escollir si estigués entre els 20 amb major capitalització.

Tot i que moltes persones coneixen l'IBEX 35 com un dels principals índexs borsaris d'Espanya, no tothom sap com es calcula exactament. Com ja hem comentat, és important tenir en compte que es basa en la capitalització de cadascuna de les empreses que el componen. És a dir, es considera el valor total de totes les accions de cada empresa que cotitza a la Borsa de Madrid.

Per assegurar la precisió del càlcul, s'utilitza un coeficient d'ajust, el qual contempla diversos factors com ampliacions de capital o les fusions entre empreses. Aquest coeficient, ajuda a evitar possibles operacions financeres que podrien influir de manera artificial a la cotització de l'índex.

En termes més específics, la fórmula per calcular-lo és la següent:

$$IBEX\ 35(t) = IBEX\ 35(t - 1) * \frac{\sum_{i=1}^{35} Cap_i(t)}{[\sum_{i=1}^{35} Cap_i(t - 1) \pm J]}$$

Fórmula 2.1: Índex IBEX 35

$$Cap_i(t) = S_i * P_i(t) \quad Cap_i(t - 1) = S_i * P_i(t - 1)$$

Fórmula 2.2: Capitalització en el moment t i t - 1

Sent:

- $IBEX\ 35(t)$ = valor de l'índex en el moment t expressat en punts d'índex.
- $IBEX\ 35(t - 1)$ = valor de l'índex en el moment $t - 1$ expressat en punts d'índex.
- t = Moment del càlcul de l'índex.
- i = Empresa inclosa en l'índex.
- S_i = Nombre d'accions computables de la companyia i pel càlcul del valor de l'índex.
- $P_i(t)$ = Preu de les accions de la companyia i inclosa en l'índex en el moment t .
- $P_i(t - 1)$ = Preu de les accions de la companyia i inclosa en l'índex en el moment $t - 1$.
- $Cap_i(t)$ = Suma de la capitalització de totes les empreses incloses en l'índex en el moment t .
- $Cap_i(t - 1)$ = Suma de la capitalització de totes les empreses incloses en l'índex en el moment $t - 1$.
- J = Quantitat utilitzada per ajustar el valor de l'índex per ampliacions de capital, etc.

La funció de la component J és assegurar que l'índex es vegi afectat per possibles operacions financeres.

Concretament, l'índex es calcula d'acord amb el de la sessió del dia anterior. Aquest valor es multiplica per un percentatge obtingut de dividir la suma actual de les capitalitzacions de les 35 empreses, entre la suma de les capitalitzacions del dia anterior. En cas d'existir alguna operació que afectés el resultat, s'afegeix en el divisor un valor de ponderació corrector.

Si volguéssim calcular la capitalització d'una empresa en específic, s'aconseguiria multiplicant el número de les accions de cada societat pel valor de cotització del dia en concret.

A continuació, es mostra el llistat de les empreses que el conformen actualment, juntament amb el seu codi identificatiu, la data d'entrada, el sector al qual pertany i, finalment, la seva ponderació a l'índex.

Taula 2.2: Empreses que conformen l'IBEX 35

| Empresa | Codi | Data d'entrada | Sector | Ponderació en % (09/2022) |
|----------------------------------|------|----------------|--|---------------------------|
| Acciona | ANA | 2015 | Construcció | 2,09 |
| Acciona Energía | ANE | 2022 | Energies renovables | 0,65 |
| Acerinox | ACX | 2015 | Mineral, metalls i transformació | 0,53 |
| ACS | ACS | 1998 | Construcció | 1,56 |
| Aena | AENA | 2015 | Transport i distribució | 3,26 |
| Amadeus IT Group | AMS | 2011 | Electrònica i programari | 5,38 |
| ArcelorMittal | MTS | 2009 | Mineral, metalls i transformació | 0,98 |
| Banco Sabadell | SAB | 2004 | Bancs i caixes d'estalvi | 0,99 |
| Banco Santander | SAN | 1992 | Bancs i caixes d'estalvi | 10,5 |
| Bankinter | BKT | 1992 | Bancs i caixes d'estalvi | 1,24 |
| BBVA | BBVA | 1992 | Bancs i caixes d'estalvi | 7,42 |
| CaixaBank | CABK | 2009 | Bancs i caixes d'estalvi | 5,37 |
| Cellnex Telecom | CLNX | 2016 | Telecomunicacions i altres | 5,6 |
| Enagás | ENG | 2003 | Electricitat i gas | 1,07 |
| Endesa | ELE | 1992 | Electricitat i gas | 1,76 |
| Ferrovial | FER | 1999 | Construcció | 4,29 |
| Fluidra | FDR | 2021 | Enginyeria i altres | 0,56 |
| Grifols | GRF | 2008 | Productes farmacèutics i biotecnologia | 1,18 |
| IAG | IAG | 2011 | Transport i distribució | 1,42 |
| Iberdrola | IBE | 1992 | Electricitat i gas | 15,85 |
| Inditex | ITX | 2001 | Tèxtil, vestit i calçat | 9,9 |
| Indra Sistemas | IDR | 1999 | Electrònica i programari | 0,34 |
| Inmobiliaria Colonial | COL | 2007 | SOCIMI | 0,58 |
| Logista | LOG | 2022 | Transport i distribució | 0,20 |
| Mapfre | MAP | 1992 | Assegurances | 0,75 |
| Meliá Hotels Internacional | MEL | 2016 | Lleure, turisme i hostaleria | 0,23 |
| Merlin Properties | MRL | 2015 | SOCIMI | 1,00 |
| Naturgy | NTGY | 2018 | Electricitat i gas | 1,17 |
| Red Eléctrica Corporación | RED | 2000 | Electricitat i gas | 2,26 |
| Repsol | REP | 1992 | Petroli | 4,10 |
| Laboratorios Rovi | ROVI | 2021 | Productes farmacèutics i biotecnologia | 0,47 |
| Sacyr | SCYR | 2022 | Construcció | 0,36 |
| Solaria Energía y Medio Ambiente | SLR | 2020 | Energies renovables | 0,48 |
| Telefónica | TEF | 1992 | Telecomunicacions i altres | 5,26 |
| Unicaja Banco | UNI | 2022 | Bancs i caixes d'estalvi | 0,25 |

(Font: Wikipedia)

Es pot veure en aquest llistat d'empreses que conformen l'IBEX-35 com hi ha molta diversitat de ponderacions. La data d'entrada més antiga és l'any 1992, i són unes quantes empreses les que van començar a formar-ne part aquell any, com per exemple, BBVA, Telefónica i Endesa. A més, podem veure que els sectors que predominen en l'índex són:

Taula 2.3: Recompte d'empreses que conformen l'IBEX 35 per sector

| Sector | Recompte |
|--|-----------------|
| Bancs i caixes d'estalvi | 6 |
| Electricitat i gas | 5 |
| Construcció | 4 |
| Transport i distribució | 3 |
| Electrònica i programari | 2 |
| Energies renovables | 2 |
| Mineral, metalls i transformació | 2 |
| SOCIMI | 2 |
| Telecomunicacions i altres | 2 |
| Productes farmacèutics i biotecnologia | 2 |
| Petroli | 1 |
| Assegurances | 1 |
| Enginyeria i altres | 1 |
| Lleure, turisme i hostaleria | 1 |
| Tèxtil, vestit i calçat | 1 |
| Total general | 35 |

(Font: Elaboració pròpia)

Si les ordenem de més ponderació a menys, obtenim que les cinc empreses que més pes tenen, és a dir, més influència sobre l'índex, són les següents:

Taula 2.4: Cinc empreses amb més ponderació a l'IBEX 35

| Empresa | Codi | Data d'entrada | Sector | Ponderació en % (09/2022) |
|------------------------|-------------|-----------------------|----------------------------|----------------------------------|
| Iberdrola | IBE | 1992 | Electricitat i gas | 15,85 |
| Banco Santander | SAN | 1992 | Bancs i caixes d'estalvi | 10,5 |
| Inditex | ITX | 2001 | Tèxtil, vestit i calçat | 9,9 |
| BBVA | BBVA | 1992 | Bancs i caixes d'estalvi | 7,42 |
| Cellnex Telecom | CLNX | 2016 | Telecomunicacions i altres | 5,6 |

(Font: Elaboració pròpia)

Podem observar que tres de les empreses que més ponderen són de les més antigues que conté l'índex, és a dir, de l'any 1992, i corresponen a Iberdrola, Banco Santander i BBVA. En aquesta taula, es torna a veure que el sector de bancs i caixes d'estalvi torna a predominar, ja que dues de les cinc empreses que més ponderen pertanyen a aquest.

2.4. Predicció de la borsa

La predicció de la borsa es refereix a l'habilitat de predir el comportament futur del mercat de valors. Es tracta d'un tema d'interès per la gran quantitat d'inversors que hi ha. L'estudi i la compressió dels moviments dels preus de la borsa de valors ha estat des de fa temps el que han estudiat els economistes. Tot i això, no és un repte gens fàcil, a causa de la variabilitat del rendiment i el seu risc.

Existeixen diversos enfocaments i tècniques per predir el comportament de la borsa. En aquest apartat, s'explicaran els diferents mètodes de predicció que actualment s'utilitzen avui dia per determinar la tendència i els possibles preus de les accions en un temps futur.

Aquestes prediccions es fan un cop s'ha elaborat l'anàlisi borsària. Aquesta anàlisi és la disciplina que estudia els moviments de les cotitzacions borsàries a través de gràfics, balanços, sèries temporals, sèries estadístiques o teories de diversos caràcters. Els tres tipus d'anàlisis borsàries més importants són:

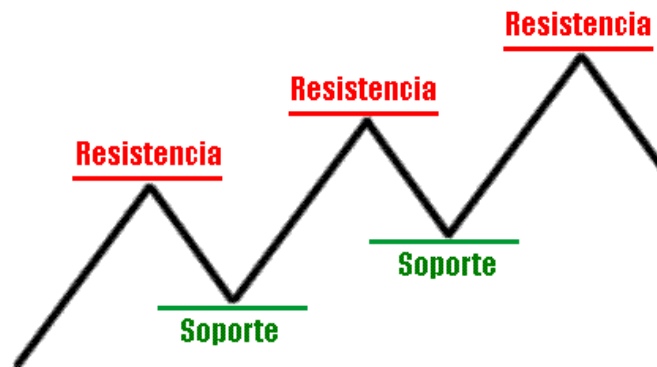


*Figura 2.3: Tipus d'anàlisi borsaris
(Font: Elaboració pròpia)*

L'anàlisi tècnica estudia els moviments borsaris mitjançant els gràfics i indicadors basats en el preu de les cotitzacions, i així poder preveure canvis en la tendència del preu de les accions. Aquest és purament matemàtic i algorítmic, és a dir, és fonamentat sempre en patrons i dades històriques. D'aquesta manera, és més útil per les prediccions a curt termini, ja que s'intenta predir el preu de l'acció en les pròximes sessions.

Es pot dividir en tres parts: l'anàlisi xartista, el qual estudia les formes dels gràfics de preus; els indicadors tècnics, que representen gràficament les fórmules basades en preus; i per últim, la teoria de les ones d'Elliot, que analitza patrons en els moviments dels preus de les accions.

Un dels principals avantatges de l'anàlisi tècnica és que pot proporcionar senyal de compra o venda d'una acció o instrument financer en temps real, com també es pot usar per identificar nivells de suport i resistència.



*Figura 2.4: Exemple de nivells de suport i resistència
(Font: Mundo Forex)*

En canvi, l'anàlisi fonamental es centra en l'estudi de factors fonamentals que afecten el valor de les accions i altres instruments financers. És a dir, estudia el preu dels actius en funció de les variables financeres que els influeixen.

L'anàlisi fonamental contempla la situació financera, la salut empresarial i el context macroeconòmic de les empreses per poder calcular quin és el preu teòric que un actiu ha de tenir. És el més utilitzat des dels principis de l'anàlisi borsària, i l'objectiu que té és avaluar l'empresa en qüestió i posar un preu aproximat del valor amb major probabilitat.

Tal com estem veient, aquest és independent als gràfics i representacions dels preus de les accions. També podem dividir-lo en dues seccions: l'anàlisi microeconòmica, on s'estudien els comptes consolidats de l'empresa, com també variables no quantitatives que poden afectar el valor; i d'altra banda, l'anàlisi macroeconòmica, on s'estudien les components pròpies de l'economia on l'empresa participa.

Un dels avantatges que té, és que es pot usar per avaluar empreses de qualsevol sector o indústria. A més, al centrar-se en factors fonamentals que impulsen el valor de les accions, pot proporcionar una visió més profunda i completa de l'empresa que l'anàlisi tècnica.

D'altra banda, l'estudi quantitatiu o, també conegut com a anàlisi quantitativa, es basa en l'estudi de sèries temporals i models matemàtics, amb l'objectiu d'analitzar els preus dels

actius financers de qualsevol mena. Aquest es tracta a part, ja que és un estudi molt ample i a més, es pot tractar en qualsevol altre tipus d'anàlisi. A més, analitza grans conjunts de dades fent servir models matemàtics. Tot i això, serà un estudi quantitatiu vàlid depenent de la qualitat de les dades proporcionades i de la precisió dels models usats.

El principal benefici d'aquesta anàlisi és la velocitat en l'estudi de grans quantitats de dades, ja que es poden processar en poc temps i obtenir resultats àgilment.

Les tècniques principals que s'utilitzen en aquest tipus d'estudi poden ser:

- Estadístiques descriptives: usades per descriure i resumir l'històric de la borsa, utilitzant els estadístics bàsics com, per exemple, la mitjana, mediana, desviació estàndard, rang i la correlació.
- Anàlisi de regressió: per poder dur a terme modelitzacions de la relació entre variables i predir el valor futur d'un actiu en funció del preu històric, el volum de la negociació, la taxa d'interès i altres factors, és a dir, variables independents.
- Anàlisi de sèries temporals: emprades per analitzar l'evolució d'una variable financera al llarg del temps, com per exemple, el preu d'un actiu, i d'aquesta manera predir el seu comportament.
- Xarxes neuronals artificials: les quals són models matemàtics que poden predir el comportament futur del mercat en funció de les dades històriques.

D'entre totes aquestes tècniques, podem trobar més específicament models com els següents:

- Models ARIMA.
- Models GARCH.
- Model de valoració d'opcions.
- Models de valoració de bons.
- Model de Markowitz.
- Arbitratge de renda fixa.
- Arbitratge de renda variable.
- Estratègies mercat-neutrals.
- Estratègies conduïdes per esdeveniments.

Un cop estudiats els diferents tipus d'anàlisi borsària tradicionals, cal destacar que l'última tendència en la predicció del mercat borsari, entre d'altres, és l'ús d'algoritmes complexos com, per exemple, les xarxes neuronals artificials, que inclouen gran quantitat de variables i dades utilitzades en els tres tipus d'anàlisi explicats anteriorment.

D'altra banda, és essencial tenir present que predir el comportament del mercat de la borsa és una feina complexa i d'alt risc. Això és degut a la gran quantitat de variables impredecibles que poden influir en el mercat. Per aquest motiu, s'aconsella comptar amb una estratègia d'inversió al llarg del temps i no confiar exclusivament en pronòstics a curt termini.

Així mateix, no es pot tenir només el criteri d'una de les anàlisis esmentades anteriorment, ja que no són fiables individualment. Per tant, és recomanable la combinació d'aquestes conjuntament, i d'aquesta manera tenir en compte altres enfocaments d'anàlisi.

3. METODOLOGIA

En aquest apartat, s'explicarà detalladament la metodologia emprada en el desenvolupament del treball, incloent-hi els mètodes de recopilació de dades, l'anàlisi de les mateixes i les tècniques emprades per arribar a les conclusions. Es descriuran totes les eleccions que s'han pres per poder dur-lo a terme, com també les decisions dels models escollits. En resum, aquest apartat és essencial per comprendre el procés de recerca i garantir la validesa i la fiabilitat dels resultats obtinguts.

3.1. Llenguatge de programació

Per poder programar els models que utilitzarem per a la predicció de l'índex IBEX 35 necessitem un programa capaç de suportar una gran quantitat de dades. El programa escollit per la realització de les tècniques de *Machine Learning* és l'RStudio.

L'RStudio és un entorn de desenvolupament integrat per l'R. L'R és un llenguatge de programació i un entorn software usat principalment per la realització d'anàlisis estadístiques i gràfics. És una eina popular entre els estadístics, analistes de dades i científics, gràcies a la possibilitat de desenvolupar programes informàtics estadístics en una àmplia varietat de camps com poden ser economia, biologia...

A més, s'ha escollit aquest programa amb motiu de la importància que ha tingut al llarg dels estudis universitaris d'Estadística a la Universitat de Barcelona conjuntament amb la Universitat Politècnica de Catalunya. El R ha estat un programa que s'ha utilitzat a la majoria d'assignatures del grau i, gràcies a això, ens ha proporcionat molts coneixements, fent que sigui l'eina que més dominem entre totes les que ens han compartit.

No obstant això, aquest treball es documenta gran part de l'assignatura cursada a quart curs d'Estadística esmentada anteriorment, Mètodes Estadístics en Minería de dades. Gràcies a aquesta, hem pogut introduir-nos en el món del *Machine Learning* i m'ha descobert un camp el qual vull continuar estudiant.

3.2. Font de les dades

Les dades de l'històric de l'IBEX 35 han estat extretes de la pàgina web *Yahoo Finance*, una font d'informació financera i econòmica que recopila una gran quantitat de dades en línia. Entre les dades que hi podem trobar s'inclouen cotitzacions d'accions, índexs borsaris, divises, matèries primeres i bons, entre altres. *Yahoo Finance* proporciona una àmplia varietat d'informació financera actualitzada en temps real, el qual la converteix en una eina útil per l'anàlisi econòmica.

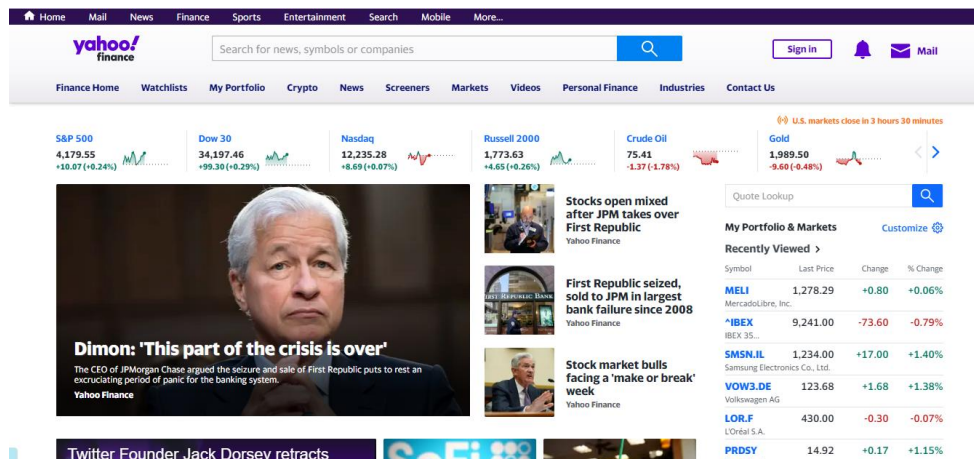


Figura 3.1: Pàgina web Yahoo Finance
(Font: Elaboració pròpia)

Un cop es realitza la recerca de l'empresa o índex del qual es vol extreure l'històric de les dades, *Yahoo Finance* ofereix escollir l'interval de dates, com també la freqüència, a triar entre diària, setmanal o mensual, juntament amb la possibilitat de descarregar-les.

Generalment, les dades es descarreguen amb un fitxer .csv, i contenen les següents columnes: Date, la data del dia; Open, el valor de l'índex en obrir el dia; High, el màxim valor que ha arribat l'índex en el dia; Low, el mínim valor que ha arribat en el dia; Close, el valor de l'índex tancar el dia; Adj Close, el valor de l'índex al tancar el dia ajustat per dividends i splits; i per últim, Volume, la quantitat de títols negociats en tot el dia.

Les dades utilitzades en aquest treball han estat els valors de l'índex IBEX 35, els quals es troben a la web *Yahoo Finance* amb l'identificador ^IBEX.

| Date | Open | High | Low | Close* | Adj Close** | Volume |
|--------------|----------|----------|----------|----------|-------------|-------------|
| Jun 23, 2023 | 9,305.60 | 9,334.10 | 9,239.70 | 9,265.80 | 9,265.80 | 124,528,300 |
| Jun 22, 2023 | 9,365.50 | 9,365.90 | 9,284.20 | 9,364.70 | 9,364.70 | 108,508,000 |
| Jun 21, 2023 | 9,431.20 | 9,463.50 | 9,415.30 | 9,436.40 | 9,436.40 | 101,629,200 |
| Jun 20, 2023 | 9,418.30 | 9,469.10 | 9,401.00 | 9,439.80 | 9,439.80 | 87,716,600 |

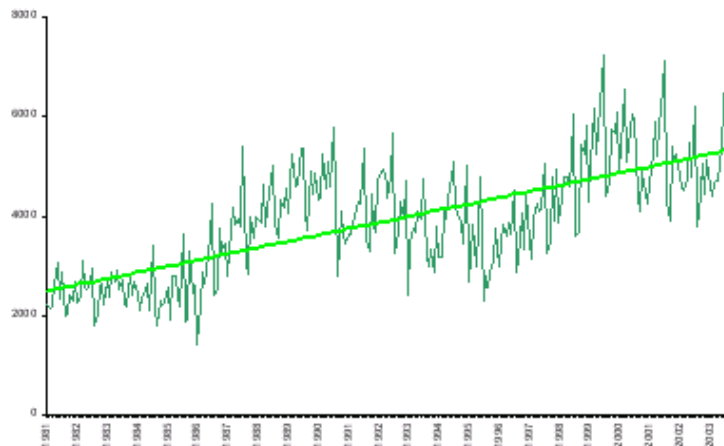
Figura 3.2: Dades de l'IBEX 35 a Yahoo Finance
(Font: Elaboració pròpia)

L'interval de temps triat ha estat des de l'01/01/2000 fins al 31/12/2022 amb freqüència diària. Per la importació de la base de dades al programa RStudio, s'ha utilitzat la funció *getsymbols*, carregant d'aquesta manera totes les variables esmentades anteriorment amb l'interval de temps i freqüència seleccionades.

3.3. Sèries temporals

El tipus de conjunt de dades que s'utilitzarà per poder dur a terme el treball final de grau és generalment conegut pel nom de sèries temporals.

Una sèrie temporal és una successió de dades o observacions mesurades en determinats moments, que fan referència a una o diverses variables i que està ordenat cronològicament (López, 2022). Aquestes, es recopilen generalment en intervals regulars i s'utilitzen per analitzar patrons i tendències al llarg del temps.



*Figura 3.3: Exemple de sèrie temporal
(Font: Universidad de Valladolid)*

Les sèries temporals són molt comunes en l'àmbit de l'economia, ja que, majoritàriament, les variables es recullen al llarg del temps per observar-hi la seva evolució, i, d'aquesta manera, no centrar-se en un valor concret d'un determinat instant.

Entorn l'anàlisi de sèries temporals, una idea clau a tenir en compte és que les observacions no són independents, és a dir, l'ordre de les dades és important.

Per dur a terme una anàlisi de sèries temporals, és vital seguir uns passos bàsics que ajudin a comprendre i modelar les dades adequadament. Alguns dels passos a seguir són els següents:

1. Visualització de les dades: és un pas primordial per començar l'anàlisi de les dades. Visualitzar les dades en forma de gràfics de línies o dispersió, pot ajudar a identificar patrons estacionals, tendències o canvis estructurals en les dades.

2. Estacionarietat: quan s'estudia una sèrie temporal és important verificar si aquesta és estacionària. Tota aquella sèrie temporal en què la mitjana i la variància siguin constants en el temps, és estacionària. En cas que no ho sigui, es poden aplicar transformacions per tornar-la estacionària.
3. Selecció del model: per poder analitzar les dades, és necessari seleccionar un model per la sèrie temporal, el qual s'adapti a les característiques d'aquesta. Els principals models que s'utilitzen són el model ARIMA (*Autoregressive Integrated Moving Average*), el model de suavitzat exponencial Holt-Winters, el model de regressió lineal, entre d'altres.
4. Estimació dels paràmetres: un cop seleccionat el model, s'estimen els paràmetres del model escollit. Es pot fer mitjançant la minimització de l'error, entre d'altres.
5. Validació del model: La validació del model és un dels passos clau per confirmar que el procés que hem seguit és correcte. Tot seguit d'estimar els paràmetres, es deu validar el model escollit per poder assegurar-nos que es podran realitzar prediccions. Per poder dur a terme aquest pas, es pot fer una divisió de les dades en un conjunt de training (entrenament) i un altre conjunt de test (prova).
6. Predicció: per últim, gràcies a que el model escollit ha estat validat, es poden realitzar prediccions sobre valors futurs de la sèrie temporal estudiada.

Aquest treball s'enfocarà inicialment a la visualització de les dades i a la seva anàlisi. Un cop realitzat això, es procedirà a la divisió de les dades en els dos conjunts esmentats anteriorment, és a dir, el conjunt d'entrenament i el conjunt de proves, per dur a terme les prediccions utilitzant els mètodes de *Machine Learning* seleccionats.

3.4. *Machine Learning*

L'objectiu principal d'aquest treball és fer una predicció dels valors de l'IBEX 35 mitjançant l'aplicació de tècniques avançades de *Machine Learning*.

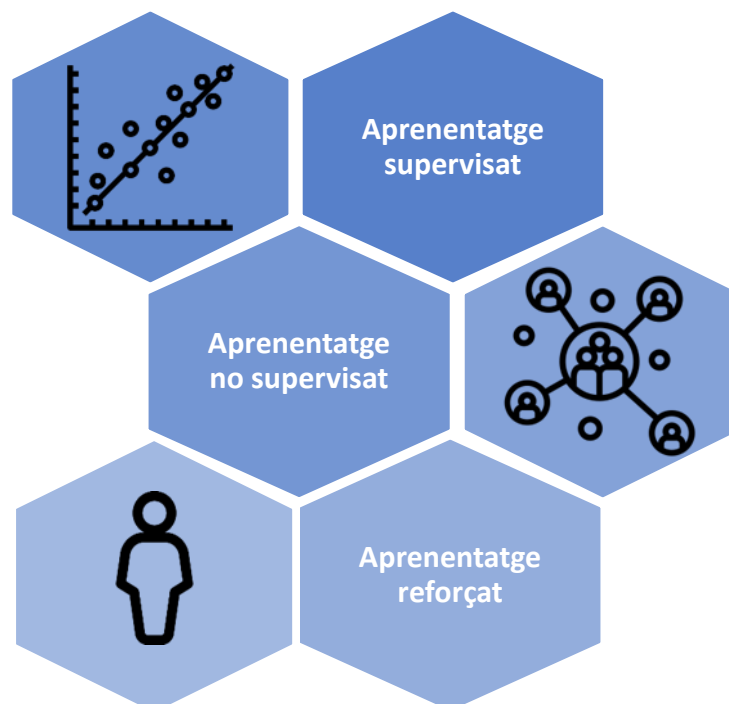
L'aprenentatge automàtic, conegut com a *Machine Learning* en anglès, és una disciplina que se centra en la comprensió i el desenvolupament de tècniques que permeten als sistemes informàtics "aprendre" mitjançant l'ús de dades, amb l'objectiu de millorar el seu rendiment en diverses tasques. Aquestes tècniques permeten a les màquines processar i analitzar grans quantitats de dades per identificar patrons, prendre decisions i adaptar-se a noves situacions sense una programació explícita per a cada cas específic (*Wikipedia contributors, 2023*).

La disciplina de l'aprenentatge automàtic, que forma part del camp de la intel·ligència artificial, empra algorismes per capacitar els ordinadors en la identificació de patrons a partir de grans quantitats de dades, cosa que els permet fer prediccions precises (anàlisi predictiva). Aquest

mètode d'aprenentatge permet als ordinadors dur a terme tasques específiques de forma autònoma, sense necessitat de programació.

Tot i que es va utilitzar per primera vegada el 1959, l'aprenentatge automàtic ha adquirit gran rellevància en anys recents gràcies a l'augment de la capacitat de processament dels ordinadors i al creixent volum de dades. De fet, les tècniques d'aprenentatge automàtic són una part essencial del camp del *Big Data*.

A continuació, s'expliquen les tres categories en què es divideixen els algoritmes de *Machine Learning*:



*Figura 3.4: Diferents tipus d'aprenentatge automàtic
(Font: Elaboració pròpia)*

L'aprenentatge supervisat són aquells algoritmes que tenen incorporat un aprenentatge previ i es basen a ensenyar a l'ordinador a través d'un sistema d'etiquetes associades a dades, les quals els permeten prendre decisions o realitzar prediccions. En aquest enfocament, es proporcionen a l'ordinador exemples d'entrades i sortides desitjades perquè pugui aprendre a relacionar-les. Amb el temps, l'ordinador serà capaç de generalitzar i predir sortides acurades per noves entrades.

Un exemple seria el detector d'*spam* de la bústia del correu electrònic. L'algoritme etiqueta el correu com *spam* depenent dels patrons que hagi après anteriorment, amb l'historial que ha

estudiat de correus. Les variables que podrien ajudar a detectar si ho és o no serien el remitent del correu, les paraules clau de l'assumpte, el contingut d'aquest, etc.

L'aprenentatge supervisat també s'utilitza en una àmplia gamma d'aplicacions, com el reconeixement de veu, la detecció del frau, la classificació d'imatges, la predicció del preu de les accions, i molts altres casos que requereixen la capacitat de predir, o, classificar dades noves en funció d'exemples previs.

Dins aquest tipus d'aprenentatge, trobem algoritmes com serien:



*Figura 3.5: Algoritmes d'aprenentatge automàtic supervisat
(Font: Elaboració pròpia)*

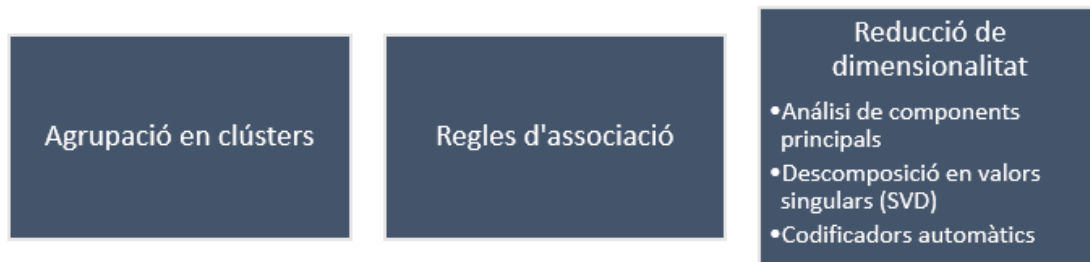
D'altra banda, tenim l'aprenentatge no supervisat, el qual implica ensenyar a l'ordinador a trobar patrons o estructures en les bases de dades sense proporcionar els exemples d'etiquetats que parlàvem just abans. En aquest cas, l'ordinador no té coneixements previs, és a dir, ha d'analitzar i trobar per ell mateix els patrons significatius i agrupar-los en categories o clústers. Es descobreixen agrupacions de dades o patrons sense necessitat de cap intervenció humana.

Per exemple, un problema de *clustering* seria proporcionar a l'ordinador un conjunt de dades sense etiquetar i demanar-li que els agrupi en grups o clústers en funció de la seva similitud.

En el camp del màrqueting s'utilitza per extreure gran quantitat de dades de les xarxes socials i usar-les per crear campanyes publicitàries molt segmentades.

Aquest tipus d'aprenentatge és fet servir generalment per la segmentació de clients, la detecció d'anomalies, la reducció de la dimensionalitat i en molts altres casos, els quals no requereixen informació prèvia.

Els diferents algoritmes que podem trobar per aquest serien els següents:



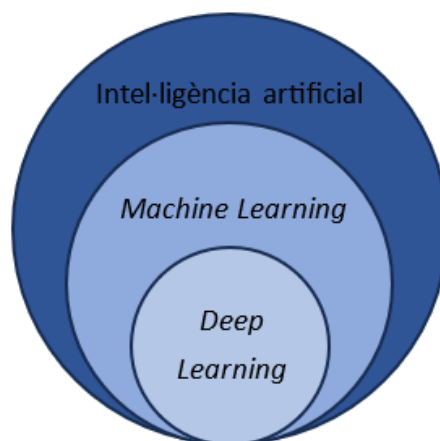
*Figura 3.6: Algorismes d'aprenentatge automàtic no supervisat
(Font: Elaboració pròpia)*

Per últim, hi trobem l'aprenentatge reforçat. Aquest, implica que un model aprengui a millorar el seu exercici en un ambient específic mitjançant la retroalimentació que rep sobre la base de les accions que realitza. L'objectiu principal és maximitzar la recompensa que s'obté mitjançant un procés de prova i error. En resum, l'aprenentatge reforçat és un mètode d'entrenament que cerca que el model aprengui per ell mateix a prendre decisions més encertades.

Aquest tipus d'aprenentatge s'utilitza per permetre el reconeixement facial, realitzar diagnòstics mèdics o, per exemple, classificar seqüències d'ADN.

Com hem esmentat anteriorment, el *Machine Learning* és una eina o enfocament utilitzat dins del camp de la intel·ligència artificial per crear sistemes o programes que puguin aprendre de manera autònoma a partir de les dades proporcionades, en lloc de requerir una programació explícita de totes les regles i decisions.

La intel·ligència artificial és una disciplina que engloba l'ús d'algorismes dissenyats amb l'objectiu de desenvolupar màquines capaces de mostrar habilitats similars a les de l'ésser humà. Aquesta tecnologia, que anteriorment semblava distant i misteriosa, s'ha convertit en part integral del nostre quotidià en els últims anys.



*Figura 3.7: Tipus d'intel·ligència artificial
(Font: Elaboració pròpia)*

Per dur a terme la predicció utilitzant tècniques de *Machine Learning*, s'ha realitzat una selecció acurada entre les diverses opcions disponibles, optant per aquelles que permeten fer regressions lineals, ja que es busca predir el valor numèric de l'índex en un moment futur basant-se en dades històriques. En aquest sentit, s'han triat tres enfocaments principals: xarxes neuronals, *random forest* i *gradient boosting*, cadascun amb les seves característiques i avantatges.

Les xarxes neuronals han estat seleccionades a causa de la seva rellevància en el camp de la mineria de dades i l'interès per explorar en més profunditat el seu funcionament i aplicacions. Aquestes xarxes, inspirades en el funcionament del cervell humà, tenen la capacitat d'aprendre i adaptar-se a patrons complexos en les dades, fet que les converteix en una eina prometedora per a la predicció de l'IBEX 35.

D'altra banda, s'ha decidit incloure-hi el mètode *random forest*. Aquesta tècnica es caracteritza per la seva capacitat per manejar conjunts de dades complexes, fer selecció de variables i manejar dades *missing* de forma eficient. En explorar *random forest*, es busca avaluar-ne el rendiment en comparació amb altres enfocaments i verificar si pot oferir una alternativa sòlida i fiable per a la predicció de l'índex.

Finalment, s'hi ha inclòs *gradient boosting*, una tècnica que fins ara no havia utilitzat. L'enfocament d'aquest es basa en la combinació de múltiples models d'aprenentatge febles per crear un model més robust i precís, cosa que ofereix l'oportunitat d'obtenir resultats prometedors en la predicció d'aquest índex.

4. MACHINE LEARNING: EINES UTILITZADES

En aquest apartat, trobarem les explicacions detallades sobre els mètodes de *Machine Learning* que s'implementaran per poder dur a terme les prediccions amb precisió i eficàcia. Tal com s'ha justificat a l'apartat anterior, les eines que s'utilitzaran són: les xarxes neuronals artificials, el *random forest* i el *gradient boosting*.

4.1. Xarxes neuronals artificials

Les xarxes neuronals, també conegudes com a xarxes neuronals artificials (ANN, Artificial Neural Network) o xarxes neuronals simulades (SNN, Simulated Neural Network), formen part de les tècniques d'aprenentatge automàtic i són fonamentals en algorismes d'aprenentatge profund. El seu nom prové del símil que té la seva estructura amb el cervell humà, imitant la forma en què les neurones biològiques es comuniquen entre si.

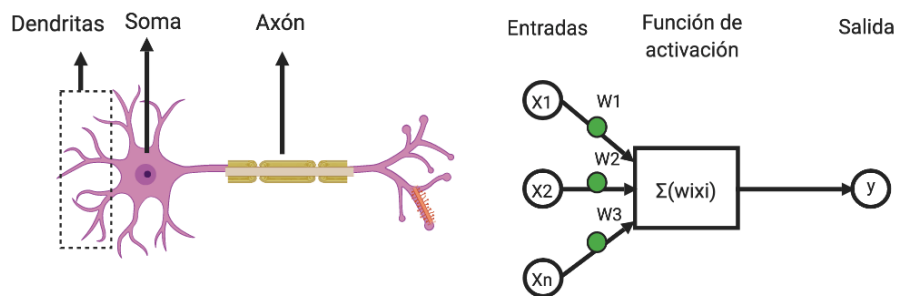
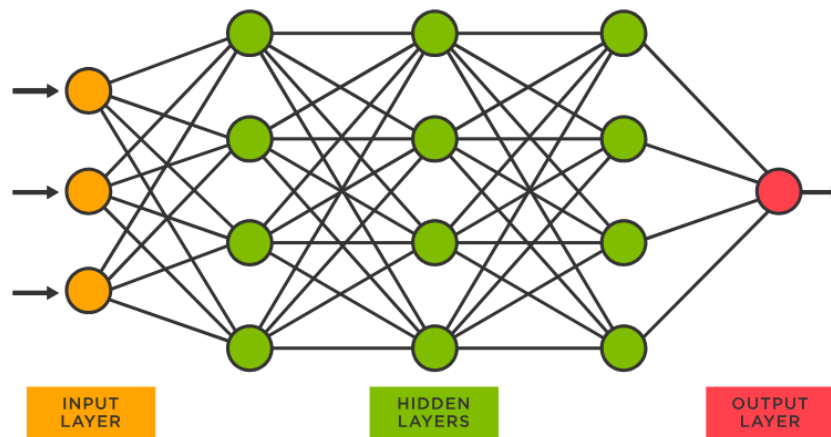


Figura 4.1: Comparativa xarxa neuronal humana i artificial
(Font: Future Lab)

Les xarxes neuronals artificials són models computacionals que consisteixen en capes interconnectades de neurones artificials, les quals processen informació i generen prediccions basades en aquesta.

Més concretament, aquestes xarxes es componen de capes de nodes que inclouen una capa d'entrada, una o diverses capes ocultes i una capa de sortida. Cada node, o en aquest cas, neurona artificial, es connecta amb altres nodes i posseeixen un pes i llindar associats. Si s'obté un valor per sobre del llindar especificat anteriorment quan hi ha la sortida d'un node individual, aquest node s'activa i envia les dades a la següent capa de la xarxa neuronal.



*Figura 4.2: Estructura d'una xarxa neuronal artificial
(Font: TIBCO)*

En aquesta imatge s'observa com la informació s'introdueix a la primera capa de neurones per l'esquerra, propagant-se cap endavant i d'aquesta manera travessant les diferents capes ocultes, fins a arribar a la capa de sortida.

L'entrenament d'una xarxa neuronal implica l'ajust dels pesos i els coeficients de les connexions entre les neurones per minimitzar l'error entre les prediccions generades per la xarxa i les dades d'entrenament.

Tot aquest procés es realitza mitjançant funcions com les següents:

- Funció de propagació: La qual relaciona els valors de les entrades, els pesos i el biaix pel càlcul de la sortida de la neurona.
- Funció d'activació: La qual determina si la neurona produirà un valor de sortida o no, és a dir, si aquesta s'activarà o no.
- Funció de transferència: Adapta el valor resultant un cop ha sortit la neurona.

També existeixen diversos tipus de xarxes neuronals artificials, i aquests es diferencien en l'estructura que tenen. Podem trobar:

- Monocapa: Aquesta, només està formada per una única capa de neurones les quals no s'interconnecten entre elles.
- Multicapa: És el tipus de xarxa més usat i el que s'utilitzarà a l'hora de definir el model de la xarxa neuronal. Es resumeix en el conjunt d'una capa d'entrada que rep valors de l'exterior, unes capes ocultes que processen la informació i, per acabar, una capa de sortida, la qual obté els resultats de la xarxa neuronal. És important recalcar que, en aquest cas, en una mateixa capa tampoc s'interconnecten. Això, només succeeix amb totes les neurones de la capa següent.

- Convolutacional: S'assemblen al tipus multicapa, tot i això, es diferencien en el fet que les neurones de cada capa no s'interconnecten amb totes de la següent, sinó només amb un subconjunt. D'aquesta manera, es redueix la quantitat d'unitats necessàries i la seva complexitat, tenint com a resultat una especialització de les neurones, és a dir, grups.
- Concurrents: En aquest tipus, trobem que les neurones no es distribueixen en capes, sinó que s'interconnecten entre elles sense seguir una estructura. Això, els permet que la informació obtinguda en iteracions prèvies, afecti el resultat del processament al llarg del temps.

D'altra banda, és important comentar que hi ha diversos tipus d'algoritmes que es poden usar per a l'entrenament i el desenvolupament de les xarxes neuronals. El que es farà servir en aquest treball és l'algoritme de *backpropagation*, un dels més comuns. Aquest, consisteix en els passos següents:

1. Assignem un valor de pes petit però diferent de zero a cada connexió entre neurones.
2. Introduïm la primera mostra del nostre conjunt d'entrenament a la capa inicial de la xarxa neuronal.
3. La informació flueix d'esquerra a dreta, activant cada neurona afectada pel pes de cada connexió, fins a assolir la capa de sortida i obtenir el resultat final per a aquesta mostra en particular.
4. Calculem l'error comès per a aquesta observació.
5. Comencem la propagació inversa, actualitzant els pesos de cada connexió de dreta a esquerra, segons la contribució de cada pes a l'error comès.
6. Es repeteixen els passos des del punt 2, actualitzant els pesos per a cada mostra o conjunt de mostres del nostre conjunt d'entrenament.
7. Un cop han passat totes les observacions del conjunt d'entrenament per la xarxa neuronal, hem completat allò que es coneix com un *Epoch* (època), els quals es poden fer tantes vegades com considerem adequades.

A continuació, trobem la fórmula generalitzada per qualsevol nombre de capes k i qualsevol neurona i dins de la capa:

$$a_i^{(k)} = f \left(u_i^{(k)} + \sum_{j=1}^{n_{k-1}} a_j^{(k-1)} w_{ji}^{(k-1)} \right), \quad i = 1 \dots n_k$$

*Fórmula 4.1: Formulació xarxes neuronals amb backpropagation
(Font: IArtificial.net)*

Un cop explicades les característiques d'aquest mètode, s'esmentaran les diferents aplicacions que tenen a la pràctica:

- Sistemes de reconeixement de veu com els emprats en els altaveus intel·ligents o *HomePod* d'Amazon, Google o Apple.
- Vehicles de conducció autònoma de Tesla i Uber.
- Els *chatbots* com Siri d'Apple, Alexa d'Amazon i Cortana de Microsoft.
- Sistemes de seguretat perimetral per a la detecció d'intrusions en temps real mitjançant el processament d'imatge captada per les càmeres de videovigilància. (Unir, 2023)

En resum, les xarxes neuronals són fonamentals en l'àmbit de l'aprenentatge automàtic i juguen un paper clau en els algoritmes de *deep learning*. Mitjançant l'entrenament amb dades, les xarxes neuronals són capaces de millorar la seva precisió amb el temps. Una vegada aquests algoritmes d'aprenentatge estan ajustats amb precisió, es converteixen en poderoses eines d'informàtica i intel·ligència artificial, permetent-nos classificar i agrupar les dades de manera eficient i ràpida.

4.2. *Random forest*

El mètode *random forest* és una tècnica d'aprenentatge automàtic. És un mètode ensemblista que combina múltiples arbres de decisió per generar prediccions precises i robustes. És un dels més utilitzats per a la seva precisió, simplicitat i flexibilitat. A més, pot usar-se tant en tasques de regressió com també de classificació.

En el mètode *random forest*, també conegut com a bosc de decisions aleatòries, es crea un conjunt d'arbres de decisió, coneguts com a "forest". Per poder entendre com funciona, és necessari explicar que és un arbre de decisió.

Un arbre de decisió en *Machine Learning* consisteix en una estructura d'un arbre, similar a un diagrama de fluxos, on cada node intern representa una característica, la branca dona lloc a la regla de decisió i els nodes finals, és a dir, les fulles, representen el resultat.

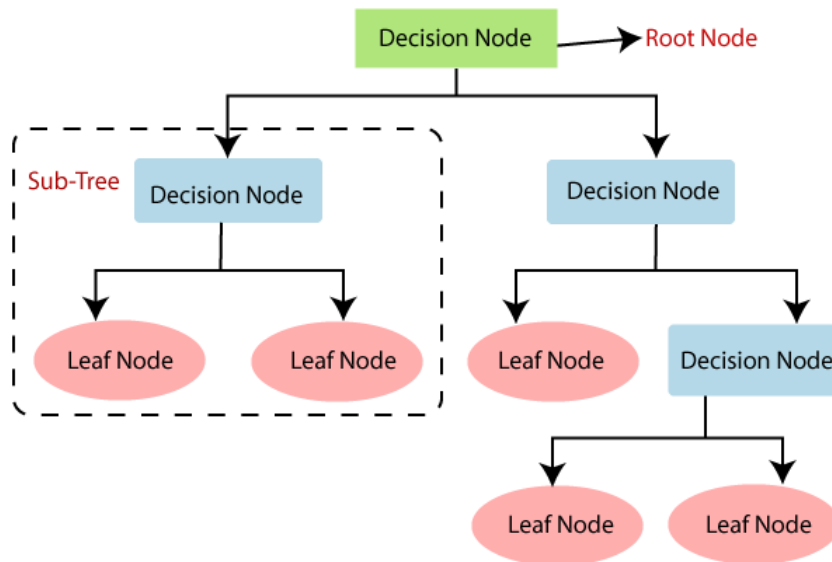


Figura 4.3: Estructura dels arbres de decisió
(Font: Javatpoint)

En el mètode de *random forest*, cada arbre és entrenat amb una mostra aleatòria de les dades disponibles i es fa servir una combinació de *bootstrap* i mostreig aleatori per garantir la diversitat de les mostres. A cada nivell de l'arbre, es realitzen divisions utilitzant les característiques més rellevants, buscant maximitzar la seva capacitat predictiva.

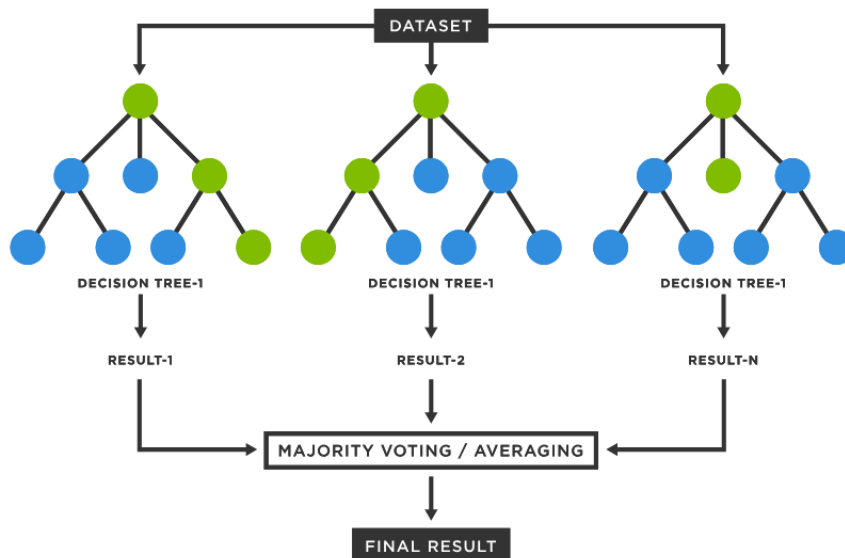


Figura 4.4: Estructura del random forest
(Font: TIBCO)

La principal diferència entre els arbres de decisió i el bosc aleatori, és que els arbres de decisió tendeixen a crear regles, i d'aquesta manera poder prendre decisions. En canvi, un bosc

aleatori escollirà les característiques aleatòriament, per poder fer observacions i d'aquesta manera, crear un bosc d'arbres de decisió on es realitzarà la mitjana dels resultats.

Els arbres construïts en el model *random forest* segueixen aquest algoritme:

1. Considerem N com el nombre total de casos de prova i M com el nombre de variables al nostre classificador.
2. Prenguem m com el nombre de variables seleccionades en un node específic per prendre una decisió. És important que m sigui molt menor que M.
3. Seleccionem un conjunt d'entrenament per construir aquest arbre i fem servir els casos de prova restants per estimar l'error.
4. A cada node de l'arbre, triem aleatòriament m variables de les quals basar la nostra decisió. Després calculem la millor divisió possible del conjunt d'entrenament utilitzant aquestes m variables seleccionades.

Quan volem fer una predicció per a un cas nou, ho passem a través de l'arbre, que baixa per les branques fins a arribar a un node terminal, on se li assigna una etiqueta. Repetim aquest procés per a cada arbre al conjunt, i l'etiqueta que més vegades es repeteixi entre els nodes terminals és la que reportem com la predicció final.

Un dels avantatges principals del mètode *random forest* és la seva capacitat per gestionar grans conjunts de dades amb múltiples variables. El model és capaç de manejar característiques numèriques i categòriques, així com de detectar i gestionar valors *missings* en les dades d'entrada. A més, el mètode *random forest* és inherentment robust a l'existència de soroll o valors atípics en les dades, ja que les prediccions finals es basen en la combinació de múltiples arbres.

Per poder entendre millor aquest mètode, s'explica a continuació diferents aplicacions que pot tenir:

- Segmentació de clients
- Pronòstic de vendes
- Autenticació de clients
- Anàlisi de comportaments del mercat

En conclusió, el mètode *random forest* és una poderosa eina d'aprenentatge automàtic per a la predicció d'índexs borsaris. Mitjançant la combinació de múltiples arbres de decisió, el model és capaç de generar prediccions precises i robustes, tenint en compte múltiples característiques i superant els desafiaments associats amb les dades borsàries, com ara la complexitat i la presència de soroll.

4.3. Gradient boosting

El *gradient boosting*, és una tècnica de *Machine learning*, concretament, d'aprenentatge automàtic, que s'utilitza per realitzar anàlisis de regressió i classificació.



Figura 4.5: Estructura del gradient boosting
(Font: AI Wiki)

El nom prové de la combinació que té entre el gradient i el *boosting*. És a dir, utilitza el *boosting* com un meta-algorisme per reduir tant el biaix com la variància, en un entorn d'aprenentatge supervisat, juntament amb l'ús de gradients per millorar el rendiment predictiu del model. Dit d'una altra manera, hi ha mètodes de *boosting*, com el *gradient boosting*, on el model es forma de manera esglaonada, permetent així optimitzar arbitràriament la funció de pèrdua diferenciable.

En el mètode de *gradient boosting*, es creen i combinen de manera seqüencial una sèrie de models d'aprenentatge febles, com ara arbres de decisió, els quals s'han explicat anteriorment, i d'aquesta manera, poder formar un model més fort. És un tipus de tècnica que implica múltiples predictors d'acord amb aquests arbres de decisió. L'objectiu principal és capturar les deficiències o residus que queden sense explicar pels models anteriors, entrenant els nous models.

A continuació, s'expliquen els passos generals de l'algorisme que usa el *gradient boosting*:

1. Selecció d'un model base feble per iniciar el procés de *boosting*, generalment un arbre de decisió poc profund.
2. S'entrena el model base amb les dades d'entrenament i es genera una predicció lineal. Inicialment, és força feble i té errors significatius.
3. Es calcula el residu, és a dir, la diferència entre el valor observat i la predicció. Aquest, representa l'error a corregir pel model.

4. S'ajusta un nou model feble als residus calculats al pas anterior, de manera que pugui predir la resta de residus o corregir els errors del model anterior.
5. S'actualitza el model afegint-hi el nou model ajustat, sumant les prediccions del nou model a les anteriors prediccions.
6. S'iteren els passos 3, 4 i 5 per crear un conjunt de models que es combinen per millorar la precisió de les prediccions. A cada iteració, es calculen els residus utilitzant la predicció actual i s'ajusta un nou model a aquests.
7. Per acabar, la predicció final s'obté sumant les prediccions de tots els models ajustats al pas anterior, el qual proporciona una predicció més acurada i robusta que la del model inicial.
8. A cada iteració es pot afegir una regularització per evitar el sobre ajustament.

A diferència d'altres models, com pot ser el *random forest*, el *gradient boosting* es centra a corregir els errors sistemàtics dels models anteriors, en comptes de recórrer a la mitjana de les seves prediccions.

Aquest tipus d'algorisme pot usar-se en el camp de l'aprenentatge de classificació. Uns dels exemples d'aplicació d'aquest, és que els motors de cerca de webs com *Yahoo* i *Yandex* utilitzen variants de *gradient boosting*.

4.4. Mètodes d'avaluació

En el context de la modelització de problemes de regressió, és fonamental avaluar l'acompliment dels models per comprendre què tan bé s'ajusten a les dades. Per aconseguir-ho, s'utilitzaran diverses mètriques que proporcionen mesures quantitatives de l'error i la qualitat de l'ajust. A continuació, es detallen les principals mètriques usades a l'avaluació de models de regressió. És important saber que aquestes tècniques per avaluar la predicció només serveixen per models de regressió, i no pas per models de classificació.

4.4.1. Mean Square Error (MSE) i Root Mean Square Error (RMSE)

L'error quadràtic mig (MSE) és una mètrica comunament utilitzada en problemes de regressió. Aquesta, s'obté calculant la mitjana dels errors al quadrat entre els valors reals i els valors predits pel model. Un cop fet això, es divideix aquesta suma entre el nombre total de mostres.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

Fórmula 4.2: Error quadràtic mig

L'MSE pondera els errors més grans de manera més significativa a causa del quadrat, cosa que pot resultar útil en escenaris on els errors grans s'han de penalitzar més que els errors petits.

D'altra banda, per poder obtenir l'arrel de l'error quadràtic mig, l'RMSE, l'únic que falta és aplicar l'arrel quadrada.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

Fórmula 4.3: Arrel de l'error quadràtic mig

Com a característica, l'RMSE té la mateixa unitat de mesura que la variable objectiu, cosa que el fa més interpretable. Utilitzant aquest tipus d'error avaluem el rendiment del model de regressió, ja que obtenim la mesura de discrepància típica entre els valors predits i els observats.

4.4.2. Mean Absolute Error (MAE) i Mean Absolute Percentage Error (MAPE)

L'error absolut mitjà (MAE) és una altra de les mètriques que s'utilitzaran per avaluar els models de regressió. Aquesta, calcula la mitjana de les diferències absolutes entre els valors reals i els valors predits.

En comparació al MSE, el MAE no eleva els errors al quadrat, i això fa que els errors més grans no tinguin tant pes. Aquest, proporciona una mesura directa de l'error mitjà i és fàcilment comprensible.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Fórmula 4.4: Error absolut mitjà

D'altra banda, es calcularà l'error percentual absolut mitjà, s'obté de la diferència percentual mitjana entre els valors reals i els valors predits, cosa que permet avaluar el rendiment del model en termes de la precisió relativa. Tot i això, és important tenir en compte que el MAPE pot tenir problemes amb valors reals propers a zero i pot donar lloc a divisions per zero.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Fórmula 4.5: Error percentual absolut mitjà

5. ESTUDI DE LES DADES

En l'àmbit financer, l'anàlisi de dades juga un paper crucial en la presa de decisions basades en evidències. El present estudi té com a objectiu endinsar-se en el complex i dinàmic entorn de l'IBEX 35, el principal índex borsari d'Espanya.

La finalitat d'aquesta investigació és examinar exhaustivament els valors que componen l'IBEX 35, amb la finalitat d'identificar patrons i tendències i, finalment, poder predir els seus valors al llarg del temps.

Com ja s'ha comentat a l'apartat anterior, per aconseguir aquest objectiu, es farà una minuciosa anàlisi d'un ampli conjunt de dades històriques recopilades durant diversos anys, del 2000 al 2022, provinents de la pàgina web *Yahoo Finance*, reconeguda en l'àmbit financer.

En el transcurs d'aquest estudi, s'empraran diverses tècniques i eines d'anàlisi de dades, com ara la visualització gràfica, càlculs estadístics i models predictius.

És important destacar que aquest estudi es centra exclusivament en l'anàlisi i no té la intenció de proporcionar recomanacions d'inversió o pronòstics precisos sobre el mercat financer. No obstant això, els descobriments i conclusions resultants podran servir com a guia i base sòlida per a aquells interessats en l'anàlisi financera i la presa de decisions relacionades amb les inversions a l'IBEX 35.

5.1. Estructura de les dades

A continuació trobem una taula amb l'explicació de la base de dades obtinguda mitjançant *Yahoo Finance*, que consta de les següents variables:

Taula 5.1: Variables de la base de dades

| Variables | Descripció | Unitats de mesura |
|------------------|--|--------------------------|
| ID | Data d'enregistrament del valor de l'IBEX 35. | Data |
| IBEX.Open | Preu d'obertura diari de l'índex IBEX 35 en el mercat financer. | Punts d'índex |
| IBEX.High | Preu màxim assolit per l'índex IBEX 35 durant el període de temps específic, el dia corresponent a la data de l'identificador. | Punts d'índex |
| IBEX.Low | Preu mínim assolit per l'índex IBEX 35 durant el període determinat, és a dir, el dia corresponent a la data de l'identificador. | Punts d'índex |

| | | |
|----------------------|--|------------------|
| IBEX.Close | Preu de tancament diari de l'índex IBEX 35. Indica el valor de l'índex en finalitzar una sessió de negociació. | Punts d'índex |
| IBEX.Volume | Volum de negociació o la quantitat total d'accions de l'IBEX 35 que es van comprar i vendre durant el període específic, el dia corresponent a la data de l'identificador. | Nombre d'accions |
| IBEX.Adjusted | Preu ajustat de l'índex IBEX 35 | Punts d'índex |

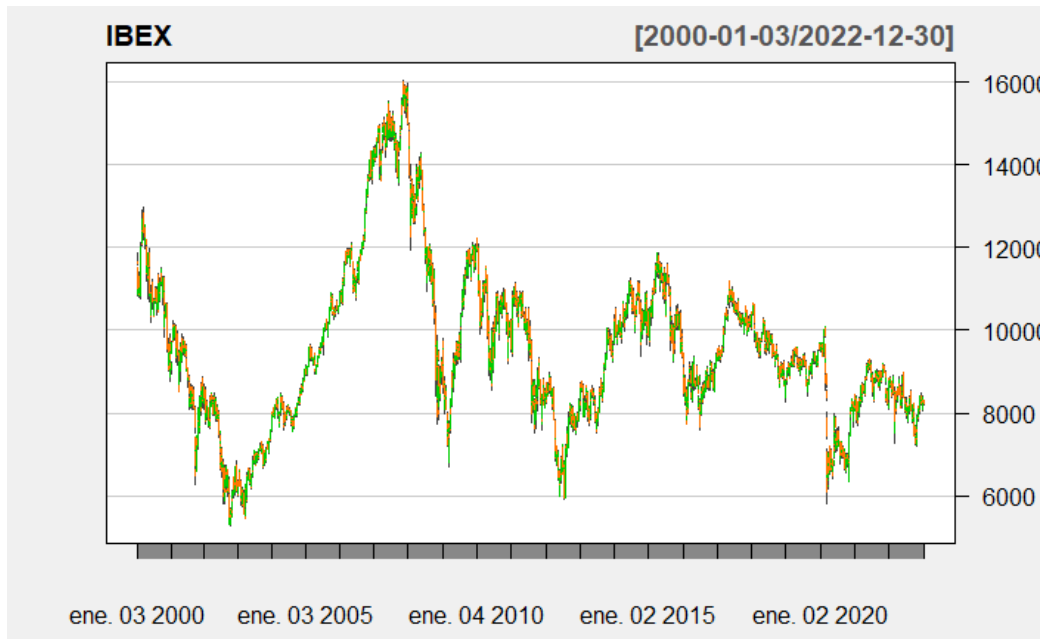
(Font: Elaboració pròpia)

Cal recordar que el preu ajustat és una mesura utilitzada per reflectir el valor real d'un actiu després de considerar esdeveniments corporatius com dividendes, desdoblaments d'accions o altres canvis en l'estructura del capital. Ajuda als inversors i analistes a tenir una visió més precisa del rendiment i la valoració dels actius al llarg del temps.

| | IBEX.Open | IBEX.High | IBEX.Low | IBEX.Close | IBEX.Volume | IBEX.Adjusted |
|------------|-----------|-----------|----------|------------|-------------|---------------|
| 2000-01-03 | 11641.4 | 11881.8 | 11574.4 | 11610.0 | 0 | 11609.99 |
| 2000-01-04 | 11530.0 | 11530.0 | 11159.8 | 11206.6 | 0 | 11206.59 |
| 2000-01-05 | 11068.1 | 11068.1 | 10824.9 | 10863.1 | 0 | 10863.09 |
| 2000-01-06 | NA | NA | NA | NA | NA | NA |
| 2000-01-07 | 10882.7 | 11137.9 | 10882.7 | 11102.4 | 0 | 11102.39 |
| 2000-01-10 | 11120.6 | 11364.3 | 11120.6 | 11173.3 | 0 | 11173.29 |
| 2000-01-11 | 11173.3 | 11234.6 | 10919.9 | 11012.4 | 0 | 11012.39 |
| 2000-01-12 | 10947.1 | 10947.1 | 10821.4 | 10851.8 | 0 | 10851.79 |
| 2000-01-13 | 10851.8 | 11169.8 | 10827.6 | 10931.9 | 0 | 10931.89 |
| 2000-01-14 | 10976.3 | 11212.1 | 10976.3 | 11184.0 | 0 | 11183.99 |
| 2000-01-17 | 11191.5 | 11394.2 | 11191.5 | 11285.6 | 0 | 11285.59 |
| 2000-01-18 | 11285.6 | 11355.4 | 11032.5 | 11066.3 | 0 | 11066.29 |
| 2000-01-19 | 11066.3 | 11134.4 | 10931.6 | 11120.7 | 0 | 11120.69 |
| 2000-01-20 | 11120.7 | 11258.1 | 10970.4 | 11047.6 | 0 | 11047.59 |
| 2000-01-21 | 11047.6 | 11073.9 | 10904.4 | 10969.3 | 0 | 10969.29 |
| 2000-01-24 | 10969.3 | 11142.2 | 10908.8 | 10955.0 | 0 | 10954.99 |
| 2000-01-25 | 10901.5 | 10901.5 | 10808.8 | 10863.5 | 0 | 10863.49 |
| 2000-01-26 | 10923.6 | 11044.3 | 10923.6 | 11031.4 | 0 | 11031.39 |
| 2000-01-27 | 11084.2 | 11293.6 | 11084.2 | 11206.6 | 0 | 11206.59 |
| 2000-01-28 | 11206.6 | 11310.9 | 10987.5 | 11009.3 | 0 | 11009.29 |

Figura 5.1: Primers registres de la base de dades importada
(Font: Elaboració pròpia)

Per tenir una idea principal de les 5.911 observacions que conté la base de dades, amb les quals es treballarà al llarg del projecte, s'ha realitzat un primer gràfic per analitzar i explorar el contingut.



*Figura 5.2: Sèrie temporal IBEX 35
(Font: Elaboració pròpia)*

Primerament, a simple vista, no podem veure cap tendència. Podem observar diversos períodes de creixement, però destaca el més prolongat i amb uns valors més alts, el qual va del 2002 al 2008, aproximadament. D'altra banda, trobem que no hi ha cap patró evident, el qual ens indicaria que les dades són aleatòries, o bé que és una sèrie temporal molt volàtil, entre d'altres. A més, podem dir que la sèrie temporal no és estacionària.

5.2. Tractament de les dades

Per poder utilitzar les dades a l'hora de realitzar models de predicció, hem hagut d'aplicar algunes transformacions. En primer lloc, s'ha modificat l'identificador de les files, el qual era la data en la qual ha tingut activitat l'IBEX 35, per una variable nova amb nom Date. D'altra banda, hem simplificat els noms amb la finalitat de simplificar-los i ser més eficients per facilitar l'escriptura del codi d'*RStudio*.

D'aquesta manera, la base de dades resulta amb els següents canvis:

| | Date | Open | High | Low | Close | Volume | Adjusted |
|----|------------|---------|---------|---------|---------|--------|----------|
| 1 | 2000-01-03 | 11641.4 | 11881.8 | 11574.4 | 11610.0 | 0 | 11609.99 |
| 2 | 2000-01-04 | 11530.0 | 11530.0 | 11159.8 | 11206.6 | 0 | 11206.59 |
| 3 | 2000-01-05 | 11068.1 | 11068.1 | 10824.9 | 10863.1 | 0 | 10863.09 |
| 4 | 2000-01-06 | NA | NA | NA | NA | NA | NA |
| 5 | 2000-01-07 | 10882.7 | 11137.9 | 10882.7 | 11102.4 | 0 | 11102.39 |
| 6 | 2000-01-10 | 11120.6 | 11364.3 | 11120.6 | 11173.3 | 0 | 11173.29 |
| 7 | 2000-01-11 | 11173.3 | 11234.6 | 10919.9 | 11012.4 | 0 | 11012.39 |
| 8 | 2000-01-12 | 10947.1 | 10947.1 | 10821.4 | 10851.8 | 0 | 10851.79 |
| 9 | 2000-01-13 | 10851.8 | 11169.8 | 10827.6 | 10931.9 | 0 | 10931.89 |
| 10 | 2000-01-14 | 10976.3 | 11212.1 | 10976.3 | 11184.0 | 0 | 11183.99 |
| 11 | 2000-01-17 | 11191.5 | 11394.2 | 11191.5 | 11285.6 | 0 | 11285.59 |
| 12 | 2000-01-18 | 11285.6 | 11355.4 | 11032.5 | 11066.3 | 0 | 11066.29 |
| 13 | 2000-01-19 | 11066.3 | 11134.4 | 10931.6 | 11120.7 | 0 | 11120.69 |
| 14 | 2000-01-20 | 11120.7 | 11258.1 | 10970.4 | 11047.6 | 0 | 11047.59 |
| 15 | 2000-01-21 | 11047.6 | 11073.9 | 10904.4 | 10969.3 | 0 | 10969.29 |
| 16 | 2000-01-24 | 10969.3 | 11142.2 | 10908.8 | 10955.0 | 0 | 10954.99 |
| 17 | 2000-01-25 | 10901.5 | 10901.5 | 10808.8 | 10863.5 | 0 | 10863.49 |
| 18 | 2000-01-26 | 10923.6 | 11044.3 | 10923.6 | 11031.4 | 0 | 11031.39 |
| 19 | 2000-01-27 | 11084.2 | 11293.6 | 11084.2 | 11206.6 | 0 | 11206.59 |
| 20 | 2000-01-28 | 11206.6 | 11310.9 | 10987.5 | 11009.3 | 0 | 11009.29 |

Figura 5.3. Primers registres de la base de dades un cop realitzat el tractament
(Font: Elaboració pròpia)

5.3. Anàlisi descriptiva

L'apartat d'anàlisi descriptiva té com a finalitat proporcionar una comprensió completa i detallada del conjunt de dades que s'estan estudiant.

A través de gràfics, taules i mesures estadístiques, es proporciona una descripció quantitativa i visual de les dades, ajudant a identificar tendències, *outliers* i altres aspectes rellevants per a l'estudi. A continuació, podem observar un resum de les variables de la base de dades:

Taula 5.2: Summary de les variables

| Variable | Mínim | 1r quartil | Mediana | Mitjana | 3r quartil | Màxim | NA |
|-----------------|-------|------------|-------------|-------------|-------------|-------------|----|
| Open | 5295 | 8283 | 9320 | 9579 | 10598 | 15999 | 64 |
| High | 5463 | 8344 | 9389 | 9651 | 10677 | 16040 | 64 |
| Low | 5267 | 8200 | 9248 | 9495 | 10517 | 15869 | 64 |
| Close | 5364 | 8281 | 9325 | 9575 | 10591 | 15946 | 64 |
| Volume | 0 | 172.200 | 124.744.700 | 121.887.446 | 214.207.550 | 975.212.000 | 64 |
| Adjusted | 5364 | 8281 | 9325 | 9575 | 10591 | 15946 | 64 |

(Font: Elaboració pròpia)

Comparant els resultats de les diferents variables de l'índex IBEX 35, podem observar que la mediana i la mitjana són similars a les variables Open, High i Close, el que indica una distribució simètrica al voltant d'aquests valors. L'abast entre el valor mínim i màxim varia entre les variables, el que indica diferències en els extrems dels valors d'obertura.

D'altra banda, els valors mínims són similars en les variables Open, Low i Close, el que indica que hi va haver moments en què l'índex va caure a nivells més baixos. Com també, els valors mínim, màxim, mediana i mitjana són similars en les variables Open, Close i Adjusted, el que indica una consistència en els preus de tancament.

També veiem com els valors de la variable Volume són significativament més alts en comparació amb les variables anteriors, el que indica una gran quantitat de transaccions realitzades durant el període considerat.

En general, en comparar les diferents variables de l'índex IBEX 35, es poden observar diferències en els extrems dels valors i en els volums de negociació, el que indica diferents nivells de volatilitat i activitat en el mercat durant el període analitzat.

A més, podem veure com hi trobem 64 valors *missings* a cada variable.

Un cop analitzats els resums de les variables, s'han graficat per poder tenir una millor visió d'aquestes.

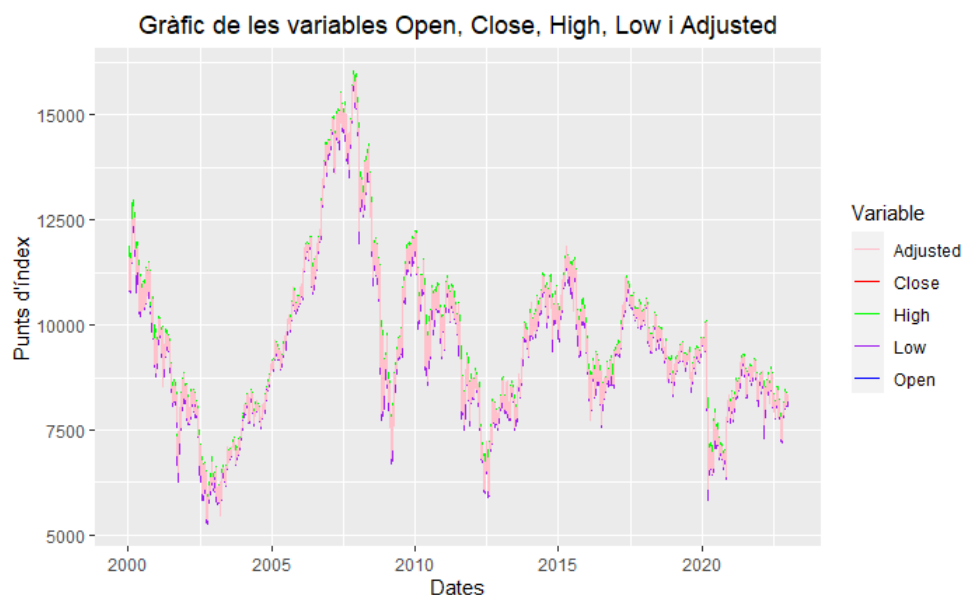
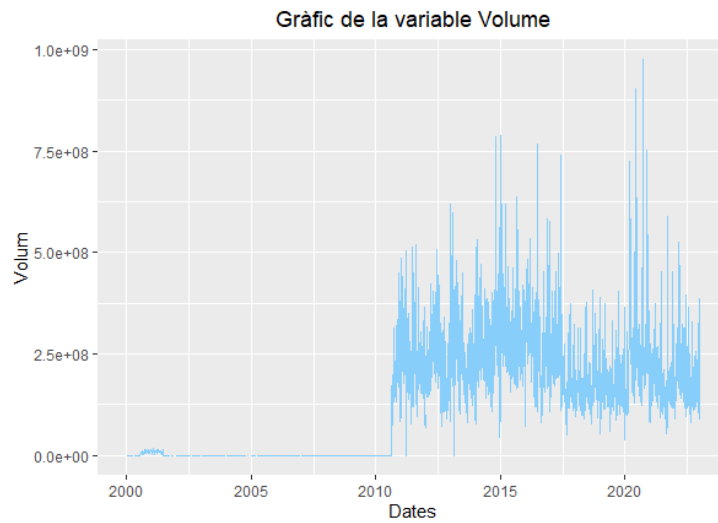


Figura 5.4: Gràfic de les variables Open, Close, High, Low i Adjusted
(Font: Elaboració pròpia)

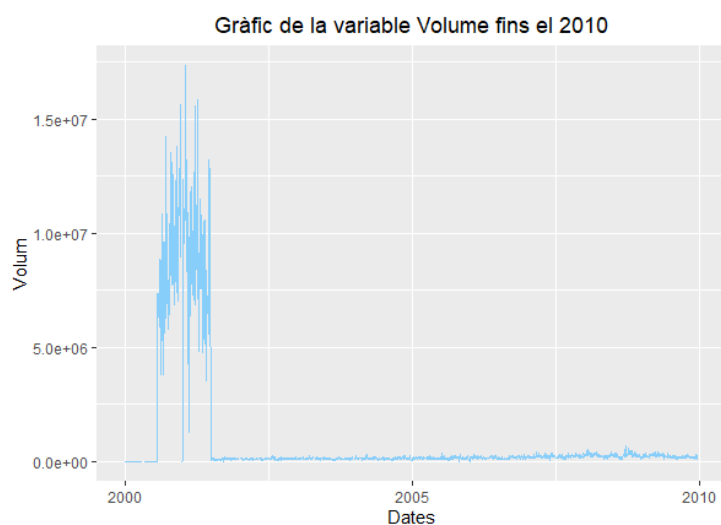
En aquest gràfic podem observar com les variables Open, Close, High, Low i Adjusted obtenen valors molt semblants, ja que pràcticament no s'aprecien els diferents colors de línies de les variables corresponents.



*Figura 5.5: Gràfic de la variable Volume
(Font: Elaboració pròpia)*

S'aprecia com hi ha un gran canvi de volum a partir del 2010 aproximadament. També es pot observar que el volum d'accions negociades més elevat s'ha aconseguit l'any 2020. Tot i això, a partir del 2010 es veu que la mitjana del volum s'aproxima a les 250.000.000 accions negociades.

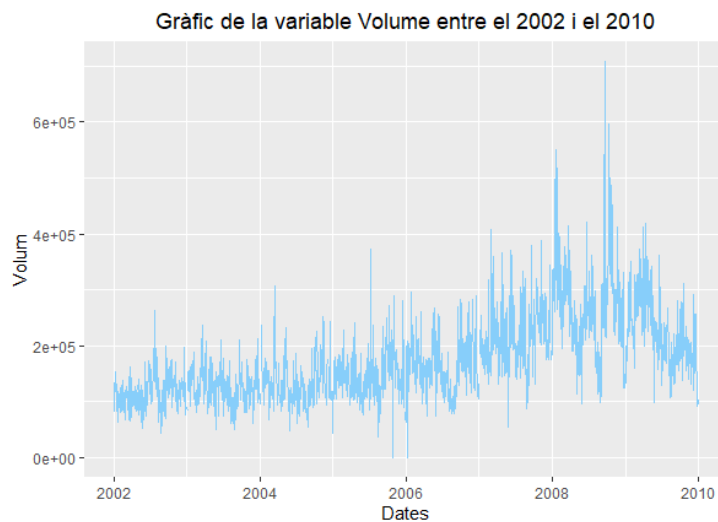
Per poder veure millor el volum que va haver-hi entre el 2000 i el 2010, s'ha realitzat un *zoom* al gràfic anterior.



*Figura 5.6: Gràfic de la variable Volume fins l'any 2010
(Font: Elaboració pròpia)*

En aquest gràfic s'observa que el volum entre el 2000 i el 2001 ronda les 10.000.000 accions negociades, menys de la meitat que en el període comentat anteriorment.

Tot i això, tornarem a fer un últim zoom entre el 2002 i el 2010 per tenir una idea de la mitjana que hi ha en gran part de les dades.



*Figura 5.7: Gràfic de la variable Volum entre el 2002 i el 2010
(Font: Elaboració pròpia)*

Ara sí que podem veure que no totes les observacions del volum són iguals a 0, sinó que són tan petites en comparació amb la resta, que s'apropen molt a 0. En aquest període, veiem com la mitjana de volum d'accions negociades és d'aproximadament 200.000.

Per tant, en general es pot afirmar que hi ha hagut diversos períodes de temps on la mitjana del volum ha experimentat canvis significatius.

Per acabar l'apartat de l'anàlisi descriptiva, s'ha fet una anàlisi bivariant de les variables, per identificar les correlacions entre elles, i s'ha obtingut el següent:

Matriu de correlació

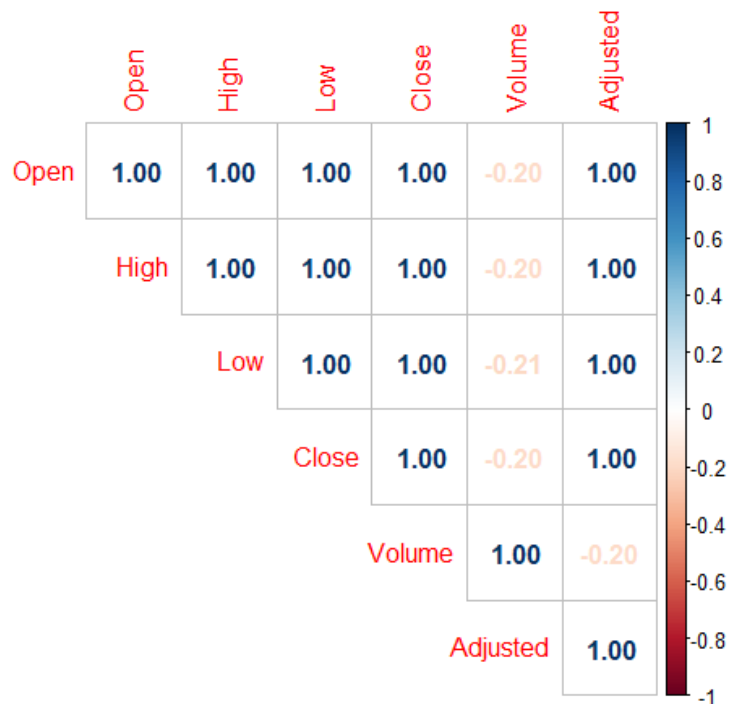


Figura 5.8: Matriu de correlació
(Font: Elaboració pròpia)

Com era d'esperar, les variables entre elles indiquen una relació lineal, perfecta i positiva. Quan la correlació és 1, ens indica que hi ha una relació directament proporcional i constant entre les dues variables, com s'observa gairebé a totes. Cada increment en una variable s'associa a un increment proporcional en l'altre, i viceversa.

L'única variable diferent torna a ser la variable Volume, que obté una correlació lineal, dèbil i negativa amb la resta de variables. El valor de correlació negativa indica que hi ha una tendència general cap a una disminució en una variable quan l'altra augmenta, però la força i la consistència d'aquesta relació són relativament baixes.

Encara i això, la variable Volume no serà utilitzada a l'hora de realitzar el pronòstic de l'IBEX 35, per tant, no serà tractada a partir d'ara.

5.4. Tractament de missings

Com hem observat anteriorment, la base de dades conté 64 *missings* per cada variable. Per poder entendre perquè consten aquests missings, s'ha volgut mirar a quines dates corresponen:

Taula 5.3: Dates en les quals es troben missings

| Any | Dates |
|-------------|--|
| 2000 | 06-01-2000, 21-04-2000, 24-04-2000, 01-05-2000, 12-10-2000, 01-11-2000, 06-12-2000, 08-12-2000, 25-12-2000, 26-12-2000 |
| 2001 | 01-01-2001, 13-04-2001, 16-04-2001, 01-05-2001, 15-08-2001, 11-09-2001, 12-10-2001, 06-12-2001, 24-12-2001, 25-12-2001, 26-12-2001, 31-12-2001 |
| 2002 | 01-01-2002, 29-03-2002, 01-04-2002, 01-05-2002, 15-08-2002, 01-11-2002, 06-12-2002, 24-12-2002, 25-12-2002, 26-12-2002, 31-12-2002 |
| 2003 | 01-01-2003, 06-01-2003, 18-04-2003, 21-04-2003, 01-05-2003, 15-08-2003, 08-12-2003, 24-12-2003, 25-12-2003, 26-12-2003, 31-12-2003 |
| 2004 | 01-01-2004, 06-01-2004, 09-04-2004, 12-04-2004, 16-08-2004, 12-10-2004, 01-11-2004, 06-12-2004, 08-12-2004, 24-12-2004, 31-12-2004 |
| 2005 | 06-01-2005, 25-03-2005, 28-03-2005, 26-12-2005 |
| 2007 | 01-01-2007, 24-12-2007 |
| 2012 | 02-01-2012 |
| 2017 | 02-01-2017 |
| 2019 | 25-12-2019 |

(Font: Elaboració pròpia)

L'IBEX 35 generalment cotitza en dies hàbils de la borsa espanyola, coneguts com a dies de negociació. No obstant això, hi ha dies en què el mercat està tancat i l'IBEX 35 no cotitza. Aquests dies poden incloure festius nacionals, festes locals i dies no hàbils específics del mercat. És per això que hi podem trobar dates a la base de dades, com pot ser el 25-12-2019. D'altra banda, poden haver-hi alguns dies on no hi ha hagut negociacions degut a suspensions del mercat o esdeveniments inusuals.

Els dies en què l'IBEX 35 no cotitza poden variar d'un any a un altre, ja que depenen dels festius i festes establerts a Espanya. Per saber amb més detall quins dies no hi tenen negociacions, seria recomanable consultar el calendari oficial de la borsa espanyola o de l'entitat reguladora corresponent per obtenir una llista actualitzada dels dies en què l'IBEX 35 no cotitza en un any determinat.

De totes maneres, aquests valors faltants poden ser estimats o imputats, per tant, s'ha decidit substituir-los per un altre valor. D'aquesta manera, garantir que tots els registres estan complets per poder-ne fer una anàlisi més exhaustiva i tenir la mateixa mida de la mostra.

Hi ha diferents tècniques per substituir els valors *missings*, en aquest cas, hem decidit usar la interpolació lineal.

La interpolació lineal és un mètode d'aproximació de valors entre dos punts coneguts en una seqüència de dades. Consisteix a estimar un valor desconegut a partir d'una línia recta que connecta els dos punts més propers a aquest valor. En la interpolació lineal, s'assumeix una relació lineal entre els punts coneguts, de manera que es calcula un valor intermedi en funció de la distància relativa d'aquest punt respecte als punts coneguts.

S'ha decidit fer aquest tipus d'interpolació perquè, veient la tipologia de dades, no hi ha canvis significatius d'un dia per l'altre, ja que aquests canvis es poden veure al llarg del temps.

Per poder dur a terme la interpolació lineal, s'ha utilitzat la funció *na.approx*.

Un cop duta a terme, tornem a realitzar un resum per assegurar-nos que no ha canviat la distribució de les dades.

Taula 5.4: Comparació mitjanes

| Variable | Mitjana anterior | Mitjana actual |
|-----------------|-------------------------|-----------------------|
| Open | 9579 | 9568 |
| High | 9651 | 9640 |
| Low | 9495 | 9484 |
| Close | 9575 | 9565 |
| Adjusted | 9575 | 9565 |

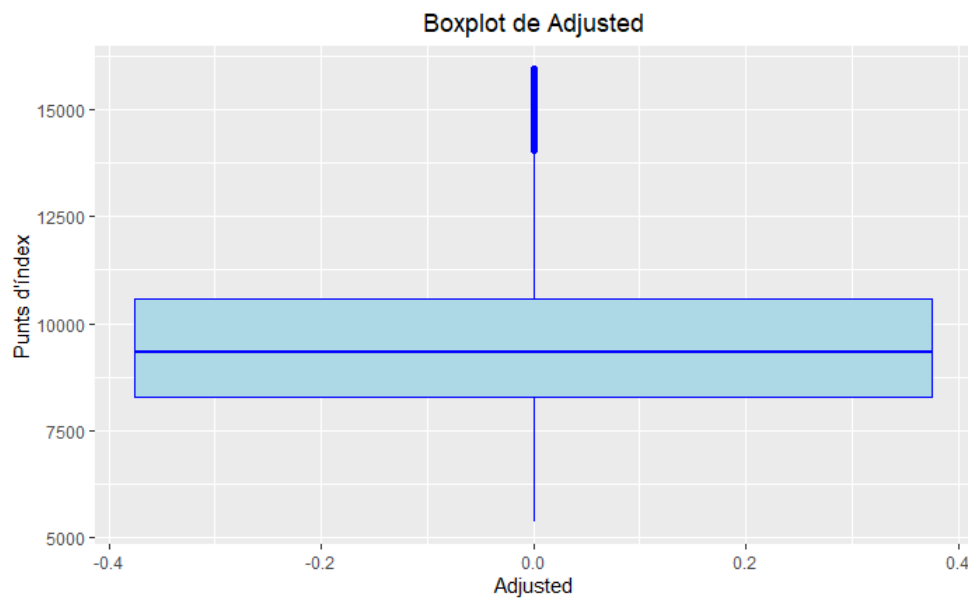
(Font: Elaboració pròpia)

Podem observar que la diferència entre la mitjana amb valors faltants i la mitjana amb aquests valors reemplaçats és molt poca en totes les variables.

5.5. Tractament d'outliers

El tractament d'*outliers* és una etapa crucial en l'anàlisi de dades, que consisteix a identificar i gestionar els valors atípics o extrems que poden afectar de manera significativa els resultats de l'anàlisi. Els *outliers* són observacions que s'allunyen considerablement del patró general de les dades i poden sorgir a causa d'errors de mesura, variabilitat natural o fins i tot esdeveniments inusuals. Aquests valors atípics poden distorsionar les estimacions estadístiques i afectar la interpretació dels resultats, per la qual cosa és fonamental abordar-los de manera adequada. En aquesta secció, explorarem diferents tècniques i enfocaments per al tractament d'*outliers*, amb l'objectiu d'obtenir un conjunt de dades més fiable i representatiu per la posterior anàlisi.

A continuació es mostra el *boxplot* de la variable Adjusted.



*Figura 5.9: Boxplot de la variable Adjusted
(Font: Elaboració pròpia)*

S'ha presentat únicament aquesta variable en concret, atès que, com s'ha observat anteriorment, les diferències entre les variables de la base de dades són molt petites, fet que fa que els *boxplots* de les altres variables siguin molt similars a aquest. Amb una atenta anàlisi, podem concloure que s'identifiquen com a *outliers* tots aquells valors que es troben aproximadament per sobre de la marca de 13.750.

Un cop vist el resultat, s'ha pres la decisió de no substituir els *outliers* identificats en el conjunt de dades. Els *outliers* representen observacions que es desvien significativament del patró general de l'índex i, tot i que poden influir en certes mesures estadístiques, també poden contenir informació valuosa o representar casos excepcionals d'interès per a l'anàlisi de l'IBEX 35. En conservar aquests valors atípics en el conjunt de dades sense realitzar cap classe de substitució, es busca mantenir la integritat i la representativitat de l'índex original, permetent que l'anàlisi posterior reflecteixi de manera precisa la naturalesa complexa i diversa dels moviments i tendències del mercat borsari.

5.6. Estadístics principals

En aquest apartat s'analitzaran diversos estadístics rellevants aplicats a les dades de l'IBEX 35. Els estadístics proporcionen mesures quantitatives que ens permeten comprendre i descriure les característiques fonamentals de la sèrie de dades de l'índex borsari. Aquests indicadors ens ajuden a obtenir informació clau sobre la distribució, tendència central, dispersió i forma de la

sèrie temporal de l'IBEX 35. En explorar els estadístics, podrem aconseguir una visió més profunda de l'evolució històrica de l'índex.

S'han realitzat diferents estadístics entre els quals trobem les mitjanes mòbils i l'índex de força relativa (RSI).

5.6.1. Mitjanes mòbils

Un dels mètodes clàssics més reconeguts i més emprats és el de les mitjanes mòbils o *Moving Averages* (MA). El càlcul de la mitjana mòbil implica agafar un conjunt de dades de preus i calcular-ne la mitjana en intervals regulars. A mesura que avança el temps, s'actualitza la mitjana mòbil eliminant el preu més antic i afegint el més recent. Això permet seguir l'evolució de la tendència sense veure's afectat pel soroll o les fluctuacions dels preus a curt termini. Aquest mètode ens aporta una idea general de la tendència.

Existeixen diversos tipus de mitjanes mòbils, com per exemple la mitjana mòbil simple o *Simple Moving Average* (SMA), la mitjana mòbil exponencial o *Exponential Moving Average* (EMA) i la mitjana mòbil Ponderada o *Weighted Moving Average* (WMA).

Mitjana Mòbil Simple

La mitjana mòbil simple (*Simple Moving Average*) consisteix a calcular la mitjana aritmètica dels n punts de dades prèvies. En aquesta tècnica bàsica de predicció, a mesura que augmenta el valor de n , s'atorga més importància a les dades històriques. Per contra, en seleccionar un valor de n més baix, es consideren més les dades més recents per fer la predicció.

$$SMA = \frac{A_1 + A_2 + A_3 + \dots + A_n}{n}$$

Fórmula 5.1: Mitjana mòbil simple

Sent A la mitjana del període escollit i n la quantitat de períodes.

A continuació, es presenten les mitjanes mòbils simples amb rangs de 30 dies (un mes), 182 dies (mig any), 365 dies (un any), 550 dies (un any i mig) i, per acabar, 1000 dies (3 anys aproximadament).

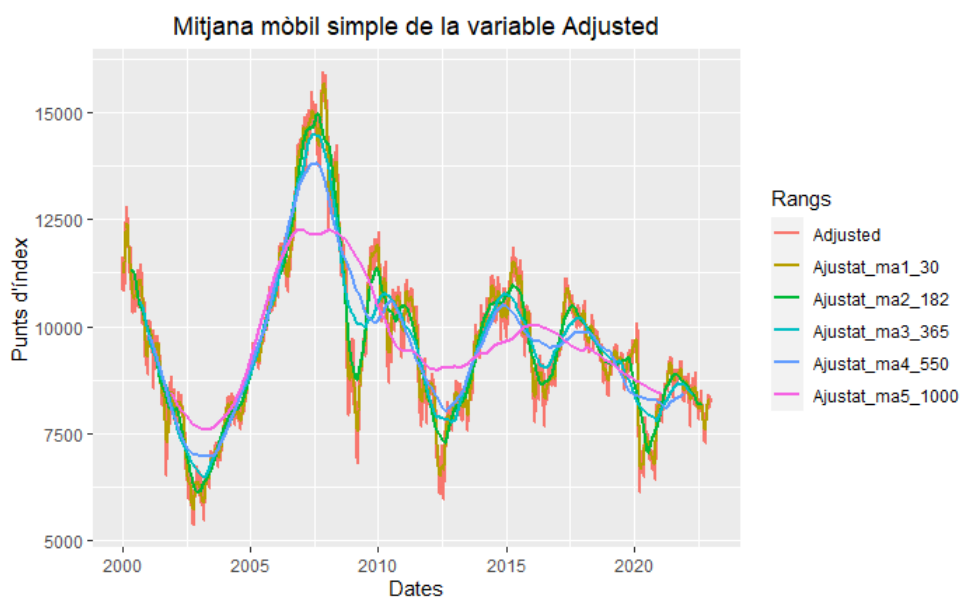


Figura 5.10: Mitjana mòbil simple de la variable Adjusted
(Font: Elaboració pròpia)

Podem observar com, a mesura que augmenta la n , més es suavitza la sèrie. La diferència és veu clarament si ens fixem en la mitjana mòbil simple amb $n=30$, la qual s'ajusta molts als valors reals, en canvi, si mirem la $n=100$, veiem que ha suavitzat de manera notable les dades analitzades.

Posteriorment, s'ha calculat els errors error quadràtic mig (MSE) i l'error percentual absolut mitjà (MAPE).

Taula 5.5: Errors de la mitjana mòbil simple

| Rangs | MSE | MAPE |
|-------------------|------------|------------|
| Ajustat_ma_1_30 | 40612.59 | 0.01611792 |
| Ajustat_ma_2_182 | 218956.95 | 0.03857670 |
| Ajustat_ma_3_365 | 438834.79 | 0.05437573 |
| Ajustat_ma_4_550 | 735848.13 | 0.07051613 |
| Ajustat_ma_5_1000 | 1572514.13 | 0.10535638 |

(Font: Elaboració pròpia)

Mitjana Mòbil Exponencial

Una EMA (*Exponential Moving Average*) o mitjana mòbil exponencial en català, és un tipus de mitjana mòbil que assigna un major pes als preus més recents i disminueix progressivament el pes dels preus més antics. A diferència de la mitjana mòbil simple (SMA), on tots els preus

tenen el mateix pes, la EMA atorga una major importància a les dades més recents, el que la fa més sensible als canvis recents en el mercat.

$$EMA = \left(Preu\ actual * \frac{2}{n - 1} \right) + \left(EMA_{anterior} * \left(1 - \frac{2}{n - 1} \right) \right)$$

Fórmula 5.2: Mitjana mòbil exponencial

A continuació, trobem les mitjanes mòbils exponencials per la nostra sèrie temporal.

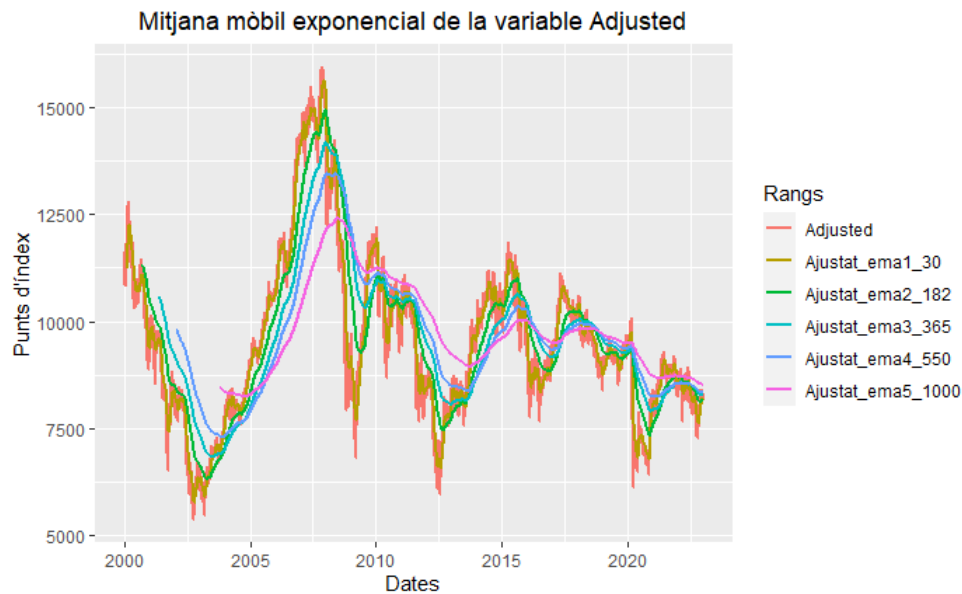


Figura 5.11: Mitjana mòbil exponencial de la variable Adjusted (Font: Elaboració pròpia)

Podem observar que passa el mateix comentat anteriorment. La diferència entre aquest tipus de mitjana mòbil i la simple, és que aquesta té les corbes inicials desplaçades lleugerament en el temps, en comparació amb les dades reals.

Els errors resultants són:

Taula 5.6: Errors de la mitjana mòbil exponencial

| Rangs | MSE | MAPE |
|---------------------------|-----------|------------|
| Ajustat_ema_1_30 | 108355.0 | 0.02627025 |
| Ajustat_ema_2_182 | 676511.6 | 0.06878842 |
| Ajustat_ema_3_365 | 1353146.9 | 0.09944421 |
| Ajustat_ema_4_550 | 1892217.0 | 0.11575965 |
| Ajustat_ema_5_1000 | 2466483.4 | 0.12070072 |

(Font: Elaboració pròpia)

Mitjana Mòbil Ponderada

A diferència de les anteriors, la mitjana mòbil ponderada (WMA) assigna pesos diferents a cada punt de dades en funció de la seva posició temporal.

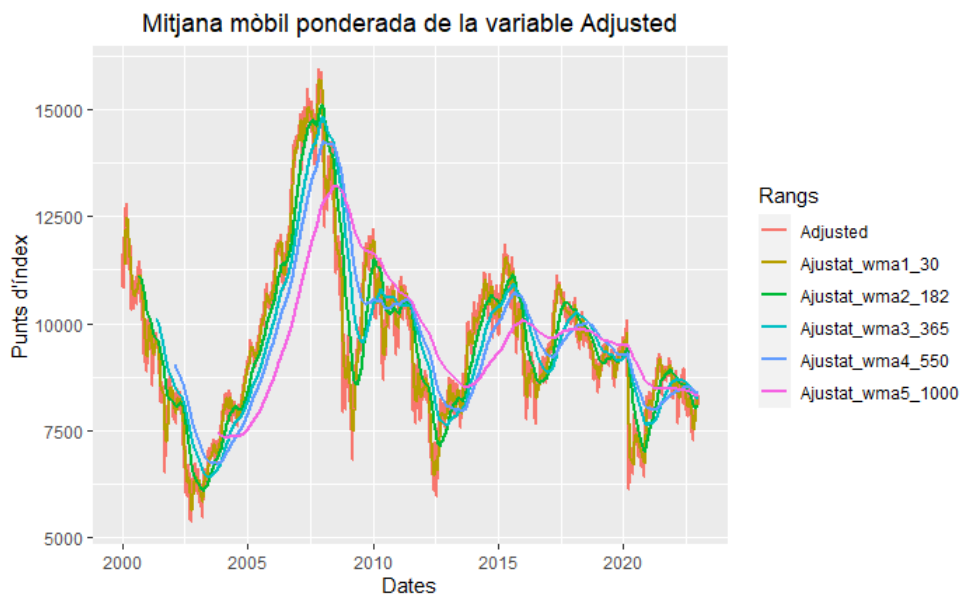
En la WMA, les dades més recents tenen un pes major, mentre que les dades més antigues tenen un pes menor. Això implica que la WMA respon més ràpidament als canvis en els preus més recents. Per calcular la WMA, es multiplica cada punt de dades per un pes corresponent i després s'obté la suma de tots els valors ponderats, dividint-la per la suma dels pesos.

$$WMA = \frac{\sum_{i=1}^N w_i * x_i}{\sum_{i=1}^N w_i}$$

Fórmula 5.3: Mitjana mòbil ponderada

Sent w_i els pesos assignats a cada x_i que són els valors de les dades.

A continuació, trobem les diferents mitjanes mòbils ponderades realitzades.



*Figura 5.12: Mitjana mòbil ponderada de la variable Adjusted
(Font: Elaboració pròpia)*

Amb aquest tipus, sembla que s'adapti millor als valors de les dades reals. Per comprovar-ho, mirarem els seus errors.

Taula 5.7: Errors de la mitjana mòbil ponderada

| Rangs | MSE | MAPE |
|--------------------|-----------|------------|
| Ajustat_wma_1_30 | 87425.4 | 0.02363720 |
| Ajustat_wma_2_182 | 526339.4 | 0.05958437 |
| Ajustat_wma_3_365 | 1110171.5 | 0.08995774 |
| Ajustat_wma_4_550 | 1641466.1 | 0.10921482 |
| Ajustat_wma_5_1000 | 2622720.9 | 0.12664153 |

(Font: Elaboració pròpia)

Per poder veure quina s'ajusta millor, s'ha realitzat una comparació entre elles que s'observa a continuació:

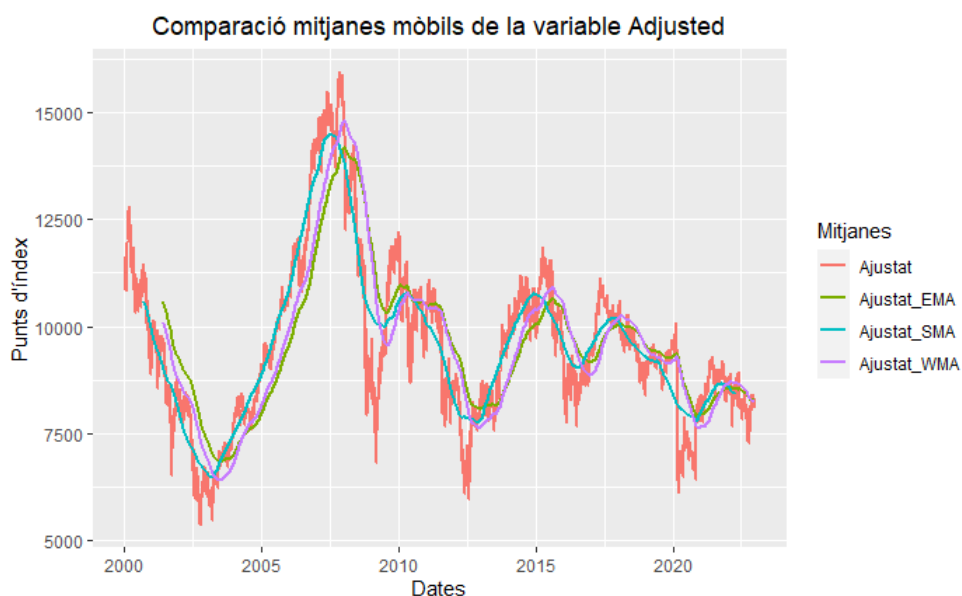


Figura 5.13: Comparació mitjanes mòbils de la variable Adjusted
(Font: Elaboració pròpia)

Taula 5.8: Comparació dels errors

| Mitjanes | MSE | MAPE |
|-------------|-----------|------------|
| Ajustat_EMA | 1353146.9 | 0.09944421 |
| Ajustat_SMA | 438834.8 | 0.05437573 |
| Ajustat_WMA | 1110171.5 | 0.08995774 |

(Font: Elaboració pròpia)

Podem observar que la mitjana mòbil que s'ajusta més al valor real de les dades és la mitjana mòbil simple, seguidament de la ponderada i, per últim, l'exponencial.

5.6.2. Índex de força relativa (RSI)

L'indicador de força relativa (RSI per les sigles en anglès, *Relative Strength Index*) és una mesura de la intensitat amb què els preus de les accions pugen o baixen al mercat borsari. Aquest indicador ofereix als inversors informació sobre la direcció del mercat i els ajuda a identificar senyals de possibles canvis a la tendència (Santander, s. f.).

Mitjançant el càlcul de la mitjana ponderada de les pujades i baixades, l'RSI proporciona una visió de la força relativa del mercat. Aquest indicador utilitza una fórmula especial per determinar els valors de l'RSI, en els quals s'obté un nombre que oscil·la entre 0 i 100. Els valors propers a 0 indiquen una situació de sobre venda, mentre que els valors propers a 100 indiquen una situació de sobre compra.

$$RSI = 100 - \frac{100}{1 + RS} \quad RS = \frac{\text{Mitjana dels preus de tancament de pujades en } n \text{ períodes}}{\text{Mitjana dels preus de tancament de baixades en } n \text{ períodes}}$$

Fórmula 5.4: RSI i RS

Els inversors utilitzen l'RSI per identificar punts d'inflexió en el mercat. Quan l'RSI està en la zona de sobre venda, és possible que el preu de l'actiu estigui infravalorat i pugui haver-hi una oportunitat per comprar. D'altra banda, quan l'RSI està en la zona de sobre compra, el preu de l'actiu pot estar sobrevalorat i es podria considerar vendre.

A continuació trobem els diferents RSI calculats per les dades de l'índex IBEX 35, en els quals s'han utilitzat els mateixos rangs que a les mitjanes mòbils.

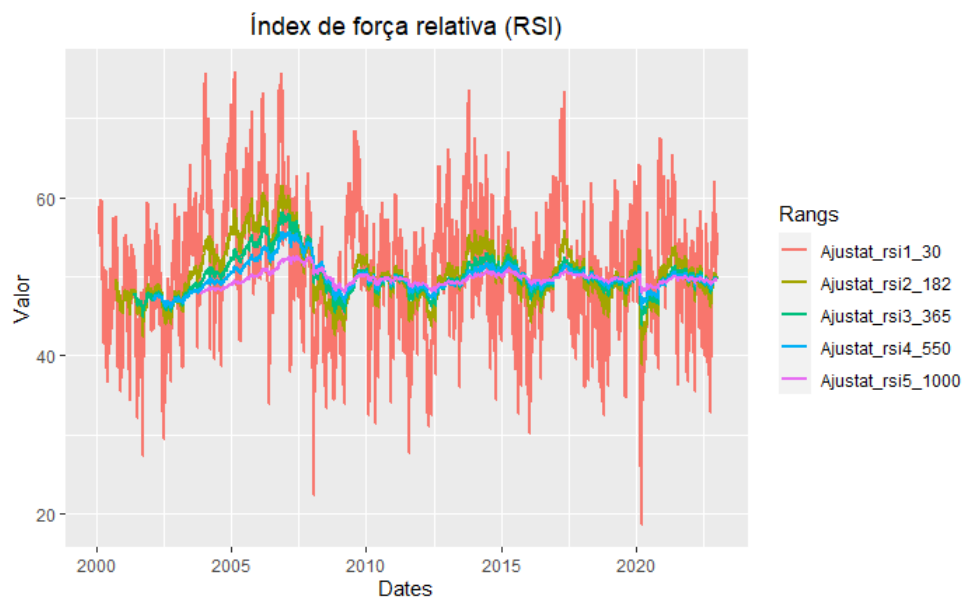


Figura 5.14: Gràfic de l'índex de força relativa (Font: Elaboració pròpia)

Podem observar com la línia n=182 és una de les millors per observar les pujades i baixades del valor de l'índex. Podem veure que en general, el valor de l'RSI està al voltant del 50, és a dir, neutre. L'any 2007 podem veure com puja, de la mateixa manera que al voltant del 2015. D'altra banda, veiem com el 2008 decau significativament, com també a principis del 2020.

En el cas de l'RSI, tornem a veure com, a més ordre té la sèrie, més es suavitza aquesta.

5.7. Preparació de les dades per utilitzar els mètodes de *Machine Learning*

Abans de començar el modelatge amb mètodes de *Machine Learning*, per obtenir resultats satisfactoris durant el procés d'entrenament i la posterior predicció, és important realitzar una adequada divisió de les dades en dos grups: dades d'entrenament i dades de prova.

Pel que fa a la divisió de les dades, hi ha dos aspectes a tenir en compte. D'una banda, si disposem de menys dades d'entrenament, les estimacions dels paràmetres tindran una major variabilitat. D'altra banda, si tenim menys dades de prova, l'estadística de rendiment mostrarà una major variabilitat. En general, es busca trobar un punt entremig on cap dels dos aspectes es superposi a l'altre.

Quan disposem d'un nombre considerable de dades, es recomana fer una divisió del 70% al 80% per a les dades d'entrenament i del 30% al 20% per a les dades de prova. En aquest cas concret, s'ha optat per una divisió del 70% per a l'entrenament i el 30% per a la prova.

6. MODELATGE

Com s'ha comentat anteriorment, per la realització de les prediccions s'ha utilitzat la variable Adjusted com a variable independent. A més, s'han inclòs altres variables per capturar la informació temporal de les dades. En aquest cas, s'han creat tres variables addicionals fent servir la variable Date: l'any, el mes i el dia.

La inclusió d'aquestes variables de data pot ser beneficiosa en l'anàlisi i el pronòstic dels preus de les accions, ja que permet capturar patrons estacionals o tendències al llarg del temps. Per exemple, és possible que hi hagi variacions estacionals en el preu de les accions relacionades amb el comportament del mercat durant certs mesos de l'any o esdeveniments econòmics específics.

Quant a les dades de prova, s'ha pres la decisió d'estendre'ls un mes addicional, específicament fins al gener de 2023. Aquesta elecció es realitza amb el propòsit d'avaluar la precisió i les característiques de les prediccions resultants en un període de temps addicional que no ha estat utilitzat en la fase d'entrenament de l'emmotllament. D'aquesta manera, es pot obtenir una comprensió millor de com es desenvolupa el model en dades futures i verificar si les prediccions són consistents amb l'evolució real del mercat durant aquest període.

6.1. Xarxes neuronals artificials

Per començar a desenvolupar el model de xarxes neuronals ha estat necessari escalar les dades. Les xarxes neuronals requereixen que les dades estiguin escalades per diverses raons com poden ser:

- La millora de la convergència, ja que si les dades tenen diferents escales o rangs, les neurones de la xarxa poden trigar més temps a ajustar els pesos adequats per capturar patrons de les dades. Un cop escalades, es redueix la variabilitat i fa que convergeixi més ràpidament durant el procés d'entrenament.
- Evitar el domini de variables: poden haver-hi variables amb un rang de valors molt més gran que altres. Si les dades no s'escalen, les observacions amb valors més grans poden dominar i tenir un impacte desproporcionat en el procés d'entrenament de la xarxa neuronal. En escalar les dades, totes les variables tenen un rang similar, cosa que permet que totes contribueixin de manera equitativa a l'aprenentatge del model.

Un cop escalades les dades, s'ha carregat les llibreries *neuralnet* i *NeuralNetTools* que seran les encarregades de proporcionar les funcions necessàries per al modelatge i la visualització de xarxes neuronals.

Per crear la xarxa neuronal s'ha utilitzat la funció *neuralnet* amb els següents paràmetres:

```
mod < -neuralnet (formula = Adjusted ~ Any + Mes + Dia,  
                data = xarxesibexsc[train,], hidden = 2, threshold = 0.01,  
                stepmax = 1e + 08, rep = 1, linear.output = TRUE)
```

Fórmula 6.1: Model utilitzat per la xarxa neuronal

El paràmetre *formula*, especifica la fórmula que defineix la relació entre les variables predictorres, Any, Mes i Dia, i la variable resposta, Adjusted. D'altra banda, li proporcionem les dades amb les quals entrenarà el model, és a dir, el conjunt de les dades escalades que pertanyen al grup *train*.

Tot seguit, trobem el paràmetre *hidden*, el qual especifica el número de capes ocultes i la quantitat de neurones a cada capa oculta. En aquest cas, s'ha establert *hidden = 2* que fa referència al nombre de capes ocultes. S'ha cregut un número adient, ja que es recomana tenir una mitjana entre la quantitat de variables d'entrada i la quantitat de variables de sortida, que en aquest cas és 3 i 1.

Finalment, trobem *threshold*, que especifica el llindar d'error mínim acceptable per la funció d'error, *stepmax* que ens demana el nombre màxim de passes d'entrenament permeses, el nombre de repeticions de l'entrenament, i *linear.output* que especifica si cal utilitzar la funció d'activació lineal a la capa de sortida de la xarxa neuronal que, en aquest cas, és TRUE ja que és apropiat pels models de regressió i així poder predir valors continus.

6.2. *Random forest*

Per la realització del model *random forest*, ha estat necessari carregar la llibreria *randomForest*. A diferència dels models de xarxes neuronals, el model *random forest* no requereix que les dades s'escalin prèviament. Això és degut a les següents raons:

- Independència entre variables: en aquest tipus de model, es construeixen múltiples arbres de decisió utilitzant subconjunts aleatoris de característiques. Cada arbre de decisió s'entrena de manera independent i, posteriorment, se'n combinen les prediccions. A causa d'aquesta independència entre variables i la naturalesa de la construcció del model, el *random forest* és menys sensible a les diferències d'escala entre les variables. No cal que les variables estiguin a la mateixa escala perquè el model *random forest* funcioni correctament.
- Selecció de divisions basada en impuresa: A cada node d'un arbre de decisió dins del *random forest*, la selecció de la variable i el punt de divisió es basa en la impuresa o el guany d'informació. L'escala de les variables no afecta directament la mesura

d'impureza utilitzada a les divisions. Per tant, no cal escalar les dades per garantir una selecció adequada de les divisions als arbres de decisió.

- Mitjana de múltiples arbres: Al *random forest*, les prediccions de múltiples arbres de decisió es combinen mitjançant mitjana o votació, depenent del tipus de problema (regressió o classificació). En combinar les prediccions de diversos arbres, es redueixen els efectes de les diferències d'escala entre les variables, cosa que fa que l'escalament previ de les dades no sigui necessari.

Per tant, el model *random forest* és robust davant de les diferències d'escala entre les variables i no requereix que les dades s'escalin prèviament.

A continuació, trobem el model utilitzat:

```
modrd <- randomForest (Adjusted ~ Any + Mes + Dia,  
                        data = xarxesibex[train,], type = "regression",  
                        ntree = 100)
```

Fórmula 6.2: Model utilitzat per al random forest

En aquest tipus de model no ha estat necessari definir tants paràmetres com en el cas anterior. Tornem a trobar la mateixa fórmula, és a dir, les variables predictores, Any, Mes i Dia, i la variable resposta, Adjusted, com també les dades a entrenar sense escalar.

D'altra banda, tenim definit el paràmetre *type*, el qual especifica el tipus de problema que s'està plantejant, és a dir, un problema de regressió que busca predir un valor numèric continu. Per últim, *ntree* ens dona la possibilitat d'establir el nombre d'arbres que es construiran en el model. En aquest cas, s'ha escollit que es creïn 100 arbres de decisió ja que és un número amb el qual el model ja pot convergir sense que es sobre ajusti.

6.3. Gradient boosting

En aquest apartat, es descriurà el procés de construcció del model *gradient boosting* i els paràmetres utilitzats per ajustar-lo. Per poder dur-lo a terme, s'ha hagut d'instal·lar les llibreries *gbm* i *caret*.

Com s'ha comentat a l'apartat de *Machine Learning*, la construcció del model *gradient boosting* implica la creació iterativa d'arbres de decisió per capturar els residus del model anterior i millorar-ne la precisió predictiva. Per aquest motiu, tampoc ha necessitat l'escalat de les dades per usar-les com a dades d'entrenament.

El model creat ha estat el següent:

```
model_gbm = gbm(Adjusted ~ Any + Mes + Dia, data = xarxesibex[train,],  
distribution = gaussian, cv.folds = 10, shrinkage = .01,  
n.minobsinnode = 10, n.trees = 100)
```

Fórmula 6.3: Model utilitzat pel gradient boosting

En aquest model trobem les variables predictores, Any, Mes i Dia, i la variable resposta, Adjusted, tot seguit de les dades d'entrenament sense escalar. També hi ha el paràmetre *distribution*, el qual especifica la distribució dels errors en el model. En aquest cas, s'assumeix que els errors seguiran una distribució normal.

D'altra banda, el paràmetre *cv.folds* indica el nombre de plec utilitzats a la validació creuada durant l'ajust del model. La validació creuada divideix el conjunt de dades en k subconjunts, on k és el nombre de plec. En aquest cas, s'usa una validació de 10 plec, és a dir, es realitzaran 10 ajustaments del model.

El paràmetre *shrinkage* fa referència a la taxa d'aprenentatge. Controla la contribució de cada arbre al model final, és a dir, un valor més petit redueix la influència de cada arbre i s'evita d'aquesta manera el sobre ajustament. En aquest cas, s'utilitza un *shrinkage* de 0.01, que indica una contribució relativament baixa de cada arbre.

Tot seguit, el paràmetre *n.minobsinnode*, estableix el nombre mínim d'observacions requerides en un node terminal de l'arbre, el que indica que si el nombre d'observacions en un node és més baix que el valor determinat, el creixement de l'arbre s'atura. En aquest cas, es requereix un mínim de 10 observacions a cada node terminal.

Per últim, el paràmetre *n.trees* indica el nombre total d'arbres a construir al model *gradient boosting*. Com més gran sigui el nombre d'arbres, més capaç serà el model per ajustar-se a les dades, augmentant d'aquesta manera el temps d'entrenament. S'ha determinat que el nombre d'arbres a construir sigui 100, per poder realitzar una comparació amb el mateix nombre d'arbres que al mètode de *random forest*.

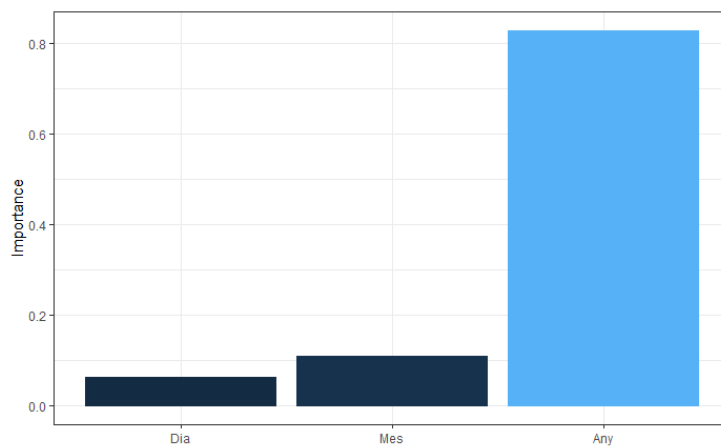
7. RESULTATS

Aquest apartat presenta i compara els resultats obtinguts mitjançant l'ús de xarxes neuronals, *random forest* i *gradient boosting* per fer prediccions en un conjunt de dades específic. S'avaluarà la precisió de cada tècnica i se n'analitzarà la capacitat per capturar relacions complexes i millorar el rendiment predictiu. Aquests resultats proporcionaran una visió detallada de les fortaleses i limitacions de cada enfocament, permetent una selecció adequada de la tècnica més apropiada per al problema en qüestió.

7.1. Xarxes neuronals artificials

A continuació trobem el gràfic amb la importància relativa de les variables predictores en el model creat amb xarxes neuronals.

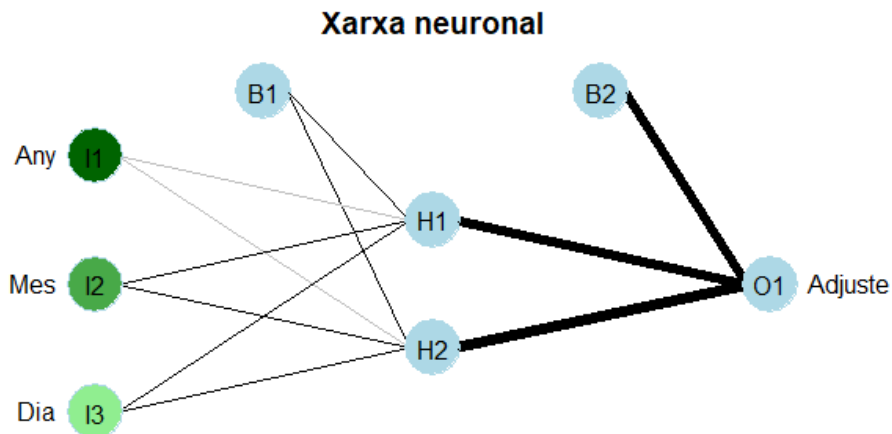
Gràfic d'importància relativa de les variables predictores



*Figura 7.1: Gràfic d'importància relativa de les variables predictores
(Font: Elaboració pròpia)*

En analitzar les dades, es pot observar clarament que la variable més rellevant és l'Any. Aquesta variable mostra un grau d'importància significativament alt en comparació de les variables Mes i Dia. De fet, el Dia és la variable que presenta la menor importància relativa en aquest context.

Podem observar la xarxa neuronal creada pel model comentat anteriorment.



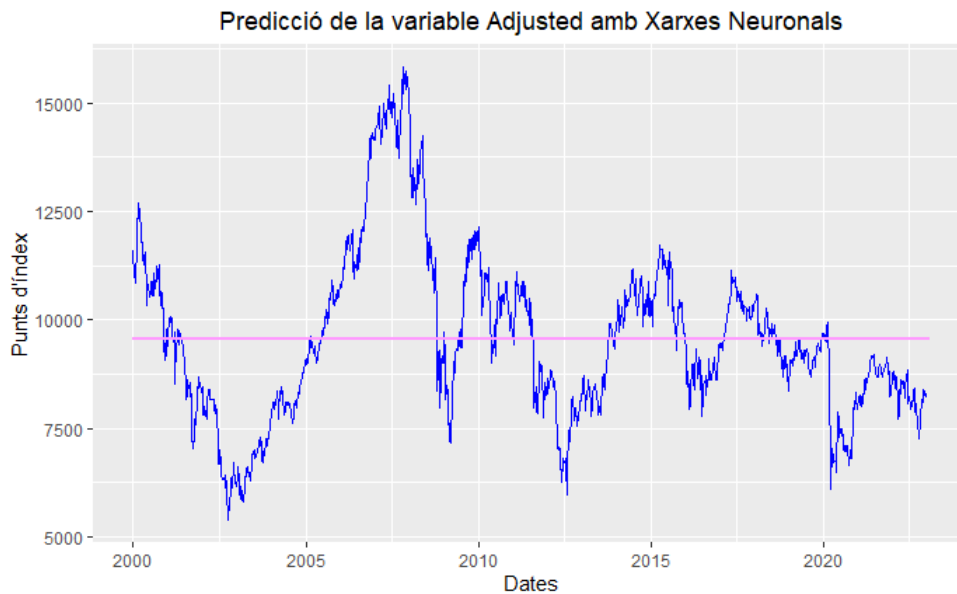
*Figura 7.2: Xarxa neuronal resultant
(Font: Elaboració pròpia)*

En observar el gràfic, es pot notar que algunes variables predictores estan representades amb un verd més fosc en comparació d'altres. Això vol dir que aquestes variables tenen un impacte més significatiu en la predicció del model i es consideren més rellevants en termes de la seva contribució per explicar la variabilitat de les dades.

D'altra banda, es presencien dues capes ocultes al model: H1 i H2. Aquestes capes ocultes són components del model de xarxa neuronal utilitzat i juguen un paper important en el processament i l'extracció de característiques de les dades.

A més, es veuen les capes de biaix, B1 i B2. Aquestes capes de biaix es poden entendre com la intersecció o punt de partida d'un model lineal. Afegir una capa de biaix permet que el model no estigui restringit a passar per l'origen i pugui ajustar millor les dades.

En el següent gràfic trobem les dades reals i la predicció resultant amb el mètode de xarxes neuronals artificials.

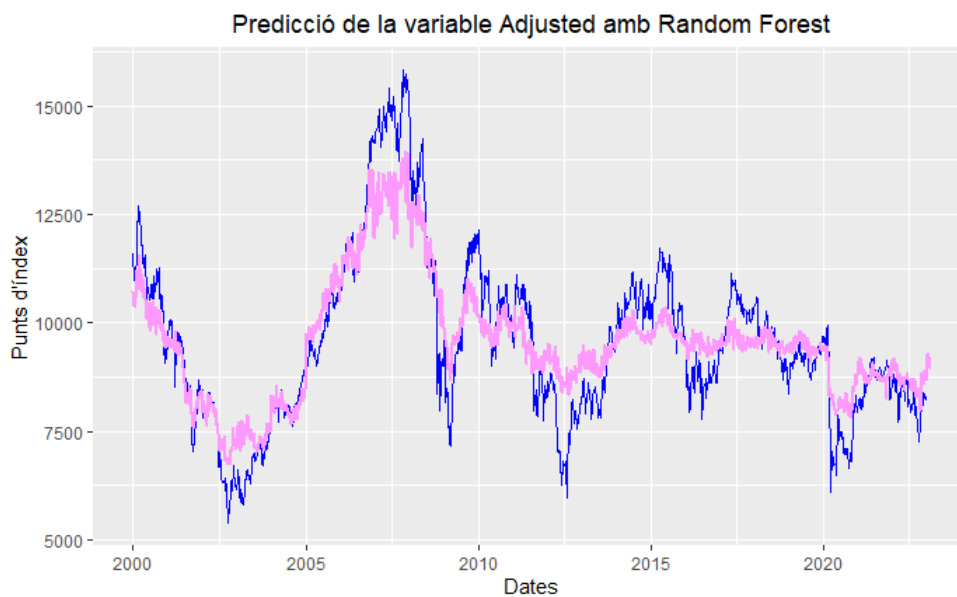


*Figura 7.3: Gràfic de la predicció de la variable Adjusted amb xarxes neuronals
(Font: Elaboració pròpia)*

Quan examinem les dades, podem notar clarament que el model no s'ha ajustat correctament a les dades originals. El resultat obtingut mostra un valor constant al llarg del temps, cosa que indica una manca de capacitat per capturar la variabilitat i els patrons presents a les dades.

7.2. *Random forest*

Pel mètode del *random forest* hem obtingut les següents prediccions:



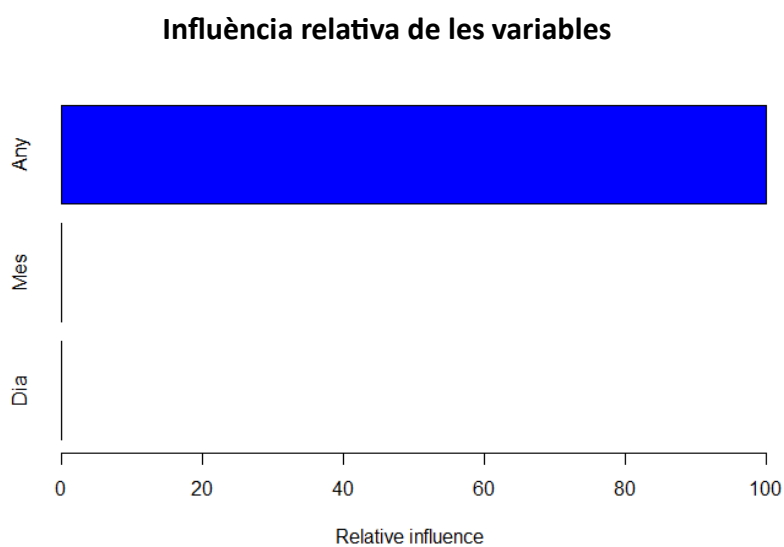
*Figura 7.4: Gràfic de la predicció de la variable Adjusted amb random forest
(Font: Elaboració pròpia)*

Es pot apreciar clarament que en aquest cas les prediccions s'acosten de manera notable als valors reals. Aquest fet ens permet intuir que el model utilitzat té una sòlida capacitat predictiva, recolzada pel mètode emprat en el procés.

El fet que les prediccions s'aproximin de manera propera als valors reals indica que el model ha capturat efectivament els patrons i les relacions subjacents a les dades. Això suggereix que les variables predictores utilitzades en el model són rellevants i estan proporcionant informació significativa per fer prediccions precises.

7.3. Gradient boosting

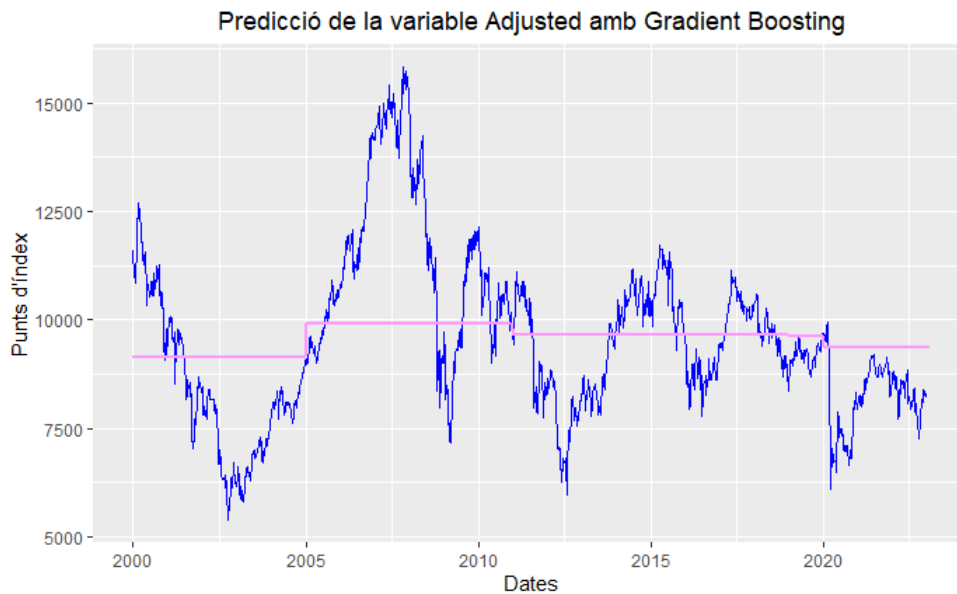
Pel mètode de *gradient boosting*, trobem a continuació la influència relativa de les variables en el model.



*Figura 7.5: Gràfic de la influència relativa de les variables
(Font: Elaboració pròpia)*

Es pot observar clarament que la variable "Any" és la que mostra una influència relativa més destacada en el model. Aquesta observació implica que l'Any té un impacte significatiu en la predicció del model en comparació amb les altres variables considerades.

D'altra banda, veiem les prediccions obtingudes, com també les dades observades en el següent gràfic.

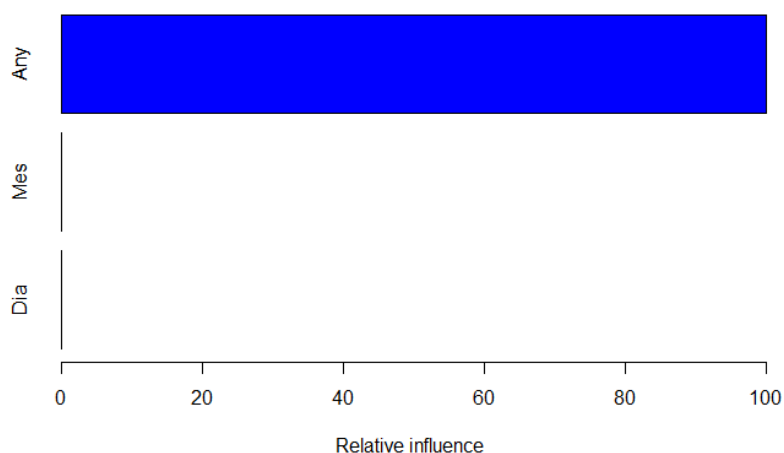


*Figura 7.6: Gràfic de la predicció de la variable Adjusted amb gradient boosting
(Font: Elaboració pròpia)*

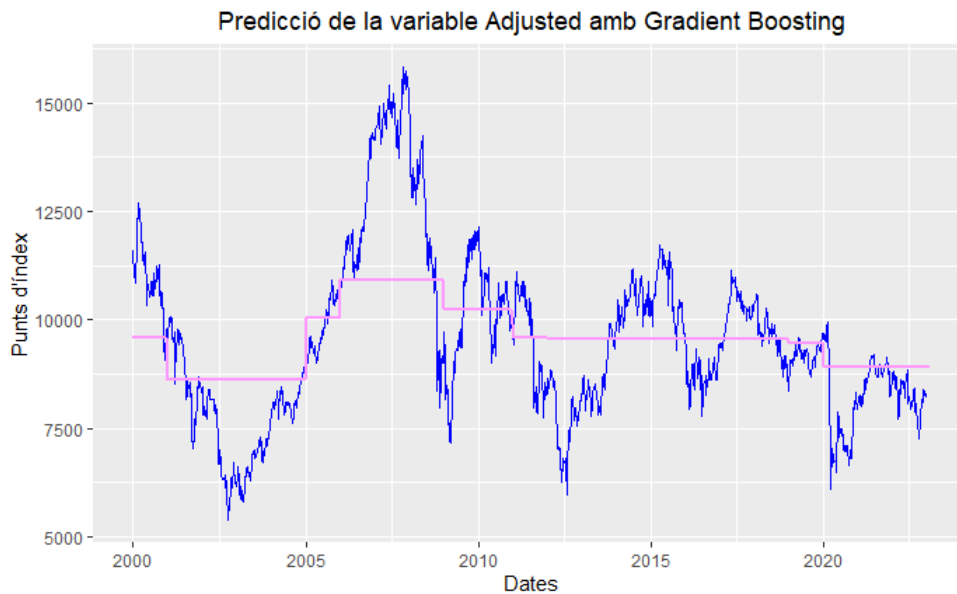
En fer una observació més detallada, podem notar que el model utilitzat s'ajusta millor en comparació del mètode de xarxes neuronals. Tot i això, malgrat aquest millor ajust, encara presenta dificultats per predir amb precisió els valors.

Per avaluar si la capacitat predictiva augmenta, s'ha decidit augmentar el nombre d'arbres al model. En aquest cas, el mateix model s'ha provat utilitzant un total de 500 arbres de decisió generats. Els resultats obtinguts són els següents:

Influència relativa de les variables



*Figura 7.7: Gràfic de la influència relativa de les variables
(Font: Elaboració pròpia)*



*Figura 7.8: Gràfic de la predicció de la variable Adjusted amb gradient boosting
(Font: Elaboració pròpia)*

Un cop més, podem observar que la influència relativa de les variables es manté constant, encara que s'aprecia un canvi a les prediccions. A primera vista, podem confirmar com, en augmentar el nombre total d'arbres de decisió, es millora la capacitat predictiva del model. D'aquesta manera, el model aconsegueix adaptar-se de manera més precisa a les dades originals, permetent reflectir de manera més efectiva les fluctuacions i canvis a l'índex que s'està analitzant.

8. AVALUACIÓ DELS RESULTATS OBTINGUTS

A continuació, es presenten els errors esmentats anteriorment per cada tècnica de *Machine Learning* utilitzada per realitzar les prediccions, és a dir, xarxes neuronals, *random forest* i *gradient boosting* amb 100 i 500 arbres.

Taula 8.1: Errors de les diferents tècniques de *Machine Learning* utilitzades

| MÈTODE | MSE | RMSE | MAE | MAPE |
|-----------------------------------|----------|----------|----------|------------|
| Xarxes neuronals | 3748867 | 1936.199 | 1487.706 | 0.1556722 |
| Random forest | 713328.4 | 844.5877 | 660.4151 | 0.06834237 |
| Gradient boosting (100 arbres) | 3226295 | 1796.189 | 1381.847 | 0.1443635 |
| Gradient boosting (500 arbres) | 2168396 | 1472.548 | 1126.732 | 0.1161151 |

(Font: Elaboració pròpia)

Entorn l'MSE, podem observar com el *random forest* i el *gradient boosting* amb 500 arbres obtenen un els valors més baixos, el que indica una millor precisió en les prediccions.

Això es pot tornar a veure amb el RMSE i MAE, els quals obtenen un menor valor en els casos de *random forest* i *gradient boosting* de 500 arbres.

Per últim, el mètode *random forest* té el valor més baix de MAPE (0.06834237), seguit pel *gradient boosting* amb 500 arbres (0.1161151). Això ens indica que, les prediccions del bosc aleatori tenen una discrepància mitjana del 6,8% respecte als valors reals. En el cas del *gradient boosting*, un 11,61% de discrepància.

També podem observar com el mètode que ha obtingut valors més alts en tots els errors és la eina de xarxes neuronals, amb valors molt propers al *gradient boosting* de 100 arbres.

Una vegada conculsa l'avaluació dels errors, és molt important destacar tant els temps de processament com el procés de preparació de les dades associades a cada mètode utilitzat.

El mètode de xarxes neuronals ha estat el que més temps ha necessitat per desenvolupar el model, al voltant dels 25 segons. En el cas d'haver especificat més capes ocultes, aquest temps hauria estat més alt. D'altra banda, el *random forest* ha estat molt més ràpid en comparació a la resta, trigant uns 2 segons. Entorn el *gradient boosting*, ha empleat 4 segons pel model de 100 arbres i 5 pel de 500 arbres.

En termes d'eficiència, és evident que el *random forest* és el mètode que obté un processament més àgil de les dades, sent també l'eina predictiva amb millor resultat.

Per últim, cal recordar que l'únic mètode que ha requerit l'escalat de dades són les xarxes neuronals, fet que suposa un pas més a l'hora de crear el model predictiu.

8.1. Selecció del millor mètode

Després d'analitzar de tots els aspectes tractats anteriorment, es constata de manera evident que el model òptim per a la realització de prediccions utilitzant les dades de l'índex IBEX 35 és el *random forest*. Aquest model ha demostrat una precisió relativament elevada, tal com s'ha avaluat mitjançant la mesura del MAPE, la qual proporciona una indicació precisa de l'exactitud de les prediccions obtingudes. A més de la seva precisió, cal destacar la seva eficiència en el processament del model. Amb un temps de construcció del model que se situa en tan sols 2 segons, és capaç de generar prediccions de manera ràpida i eficaç. Aquesta eficiència és de gran importància en el context financer, on les decisions veloces basades en prediccions acurades són essencials per a assolir l'èxit desitjat.

A més de la precisió i l'eficiència mencionades, el *random forest* es distingeix per la seva facilitat d'aplicació. Aquesta tècnica d'aprenentatge automàtic s'adapta a diverses configuracions i pot ser utilitzada amb gran encert en un ampli ventall de problemàtiques relacionades amb la predicció. La seva estructura modular i la seva capacitat per a gestionar dades amb una elevada dimensionalitat fan d'aquesta opció una elecció versàtil per als professionals dels àmbits financers i altres àrees que treballen amb dades complexes.

En resum, el *random forest* s'ha consolidat com l'elecció òptima per a la realització de prediccions utilitzant les dades de l'índex IBEX 35. Aquesta elecció robusta proporciona una base sòlida per a la presa de decisions informades i pot contribuir a obtenir resultats positius en el context del mercat financer.

9. CONCLUSIONS

Per conèixer l'evolució de l'índex IBEX 35, ha estat necessari posar-se en context sobre que és aquest índex i tot el seu entorn com la borsa de valors i els mercats financers. Aquest fet ha donat pas a assolir el primer objectiu proposat. S'ha realitzat una recerca exhaustiva d'informació per tal d'assolir el coneixement necessari per dur a terme el següent objectiu.

Per poder dur a terme l'anàlisi de l'IBEX 35, s'ha realitzat l'extracció de dades de la pàgina web *Yahoo Finance*, on s'ha escollit el període de temps i la freqüència de les observacions, en aquest cas, del 2000 al 2022 i amb freqüència diària.

En aquest anàlisi, s'ha contrastat que les variables que conté la base de dades tenien distribucions similars, sent les dades molt semblant en cada variable. Les dades faltants han estat interpolades i hem vist que les mitjanes no es desviaven dels valors originals.

A l'hora de realitzar les mitjanes mòbils i l'índex de força relativa, s'ha usat la variable *Adjusted*, ja que contempla ajustos per factors externs i elimina l'efecte d'esdeveniments corporatius. S'ha observat que la mitjana que millor s'ajusta a les dades és la mitjana mòbil simple, seguida de la ponderada i l'exponencial. A més, s'ha contrastat que una selecció menor de n observacions per realitzar mitjanes mòbils, és a dir, períodes més curts de temps, obté millors resultats. Entorn l'índex de força relativa, s'ha comprovat que succeeix el mateix fenomen.

Tot seguit, s'ha aconseguit un alt nivell de comprensió i domini dels conceptes, algorismes i enfocaments utilitzats en l'àrea de l'aprenentatge automàtic, especialment en el context de la predicció de l'IBEX 35.

Per poder dur a terme les prediccions amb tècniques de *Machine Learning*, s'han dividit les dades en dos grups. El 70% d'aquestes s'han utilitzat per l'entrenament dels models i el 30% restant s'ha usat per testejar. Un cop fet, s'han implementat els tres tipus diferents de mètodes de *Machine Learning*: xarxes neuronals artificials, *random forest* i *gradient boosting*. A través de l'aplicació d'aquestes eines, s'ha realitzat la comparativa entre els resultats obtinguts.

Entorn les xarxes neuronals, ha estat necessari aplicar l'escalat de les dades. S'ha obtingut que la variable Any és la que té més importància relativa, pel fet que pot contenir informació crucial relacionada amb patrons estacionals o tendències a llarg termini que influeixen en les dades. D'altra banda, la predicció no ha estat del tot favorable, ja que no s'ha adaptat a la sèrie original. A més, el temps de processament del model ha estat elevat.

El *random forest* ha estat el millor model implementat. Ha obtingut resultats positius, adaptant-se a la sèrie amb els seus pics i baixades, i amb un MAPE resultant de 0.068, bastant

més la meitat de la resta dels models. Cal destacar la rapidesa de processament del model i la seva facilitat d'aplicació, ja que no ha calgut escalar les dades.

El *gradient boosting*, última tècnica implementada, ha estat usat dos cops. Primer s'ha executat amb 100 arbres de decisió per comparar amb el *random forest*, que ha utilitzat els mateixos. Tot i això, ha obtingut pitjors resultats. Per aquest motiu, s'ha decidit incrementar el nombre d'arbres a 500 per veure les diferències, i s'ha obtingut millor resultat que l'anterior i que les Xarxes Neuronals, però pitjor que el *random forest*. En aquesta tècnica tampoc ha estat necessari escalar les dades.

En virtut d'aquesta anàlisi comparativa, s'ha identificat i seleccionat el mètode o algorisme de *Machine Learning* que ha demostrat ser el més adequat i eficient per a la predicció de l'IBEX 35. Després de l'estudi dels resultats, s'ha constatat que el model de *random forest* ha assolit el millor rendiment en termes de precisió i estabilitat.

Cal destacar que predir amb exactitud l'índex IBEX 35 presenta un desafiament considerable. Aquest índex està influït per una pluralitat de factors complexos i interrelacionats, com ara els esdeveniments econòmics, les polítiques governamentals, els canvis en el clima empresarial i les tendències del mercat global. A més, l'entorn econòmic i financer és inherentment volàtil i susceptible a canvis imprevisibles. Això converteix la predicció precisa del comportament de l'IBEX 35 en una tasca extremadament difícil. Malgrat els avenços en les tècniques de l'aprenentatge automàtic, és imprescindible mantenir una perspectiva realista i reconèixer els límits de la predicció en aquest context específic. Les prediccions sempre estan subjectes a incertesa, i cal tenir en compte altres factors i anàlisis addicionals per prendre decisions informades en el camp de les inversions financeres.

10. BIBLIOGRAFIA

Acción - ¿Qué es una acción en una empresa? | SumUp Facturas. (s. f.). SumUp - Una forma más inteligente de cobrar. <https://www.sumup.com/es-es/facturas/glosario/accion/>

colaboradores de Wikipedia. (2022). Mercado financiero. Wikipedia, la enciclopedia libre. https://es.wikipedia.org/wiki/Mercado_financiero

colaboradores de Wikipedia. (2023a). Yahoo! Finance. Wikipedia, la enciclopedia libre. https://es.wikipedia.org/wiki/Yahoo!_Finance

colaboradores de Wikipedia. (2023b). Media móvil. Wikipedia, la enciclopedia libre. https://es.wikipedia.org/wiki/Media_m%C3%B3vil

colaboradores de Wikipedia. (2023c). IBEX 35. Wikipedia, la enciclopedia libre. https://es.wikipedia.org/wiki/IBEX_35

colaboradores de Wikipedia. (2023d). Boosting. Wikipedia, la enciclopedia libre. <https://es.wikipedia.org/wiki/Boosting>

Cómo funcionan las predicciones en la Bolsa de Valores. (s. f.). <https://www.mejorbrokerdebolsa.com/predecir-el-precio-en-bolsa-con-garantias/>

Corporativa, I. (s. f.-a). Discover the main benefits of Machine Learning. Iberdrola. <https://www.iberdrola.com/innovation/machine-learning-automatic-learning>

Corporativa, I. (s. f.-b). ¿Somos conscientes de los retos y principales aplicaciones de la Inteligencia Artificial? Iberdrola. <https://www.iberdrola.com/innovacion/que-es-inteligencia-artificial>

DO Analytics. (2021a, agosto 9). Pronosticar precios de acciones con Machine Learning en R (Parte 1) [Vídeo]. YouTube. <https://www.youtube.com/watch?v=5Z7bfs6WJIo>

DO Analytics. (2021b, agosto 11). Pronosticar precios de acciones con Machine Learning en R (Parte 2) [Vídeo]. YouTube. https://www.youtube.com/watch?v=C_3ojnjLUMQ

DO Analytics. (2021c, agosto 15). Pronosticar precios de acciones con Machine Learning en R (Parte 3) [Vídeo]. YouTube. https://www.youtube.com/watch?v=RBr3yRdE_LA

DO Analytics. (2021d, agosto 20). Pronosticar precios de acciones con Machine Learning en R (Parte 4) [Vídeo]. YouTube. <https://www.youtube.com/watch?v=bo3UfkOE6Hk>

El IBEX 35 cumple 30 años y se consolida como marca líder de la economía y las empresas españolas | BME Bolsas y Mercados Españoles. (s. f.). Copyright © BME 2023. https://www.bolsasymercados.es/esp/Sala-Comunicacion/Nota-Prensa/20220113/nota_20220113_1/El_IBEX_35_cumple_30_anos_y_se_consolida_como_marca_lider_de_la_economia_y_las_empresas_espanolas

García, J. D. V. (2020, 18 septiembre). Redes neuronales desde cero (II): algo de matemáticas - IArtificial.net. IArtificial.net. https://www.iartificial.net/redes-neuronales-desde-cero-ii-algo-de-matematicas/#Problema_general_de_la_asignacion_de_pesos

Gradient Boosting - AI Wiki. (s. f.). <https://machine-learning.paperspace.com/wiki/gradient-boosting>

Guimerá, Martín. (2020). Predicción de índices bursátiles por medio de redes neuronales artificiales. Aplicación al caso del IBEX-35.

Heredia, Nicolás. (2016). Predicción del precio de acciones mediante técnicas de minería de datos.

Historia BME. (s. f.). Bolsas y Mercados Españoles. https://www.bolsasymercados.es/bme-exchange/docs/docsSubidos/Webs-Bolsas/Historia_BME.pdf

How to apply gradient boosting in R for regression -. (2022, 22 diciembre). ProjectPro. https://www.projectpro.io/recipes/apply-gradient-boosting-r-for-regression#mcetoc_1g4gaf4r2d

Index IBEX 35 | BME Exchange. (s. f.). BME Exchange. <https://www.bolsasymercados.es/bme-exchange/en/Indices/Ibex>

Index of /tutorial/machine-learning. (s. f.). <https://static.javatpoint.com/tutorial/machine-learning/>

Introducción a las Redes Neuronales Pt. I. (2019, 25 junio). Future Lab. <https://futurelab.mx/redes%20neuronales/inteligencia%20artificial/2019/06/25/intro-a-redes-neuronales-pt-1/>

Introduction to Random Forest in Machine Learning. (s. f.). Engineering Education (EngEd) Program | Section. <https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/>

List of Stock Markets | TradingHours.com. (s. f.). TradingHours.com. <https://www.tradinghours.com/markets>

López, J. F. (2022a). Análisis bursátil. Economipedia. <https://economipedia.com/definiciones/analisis-bursatil.html#:~:text=El%20an%C3%A1lisis%20burs%C3%A1til%20es%20la,a%20un%20campo%20muy%20amplio.>

López, J. F. (2022b). Serie temporal. Economipedia. <https://economipedia.com/definiciones/serie-temporal.html>

Mundoforex. (2023). Soportes y Resistencias. Mundo-Forex.com. <https://www.mundoforex.com/soportes-y-resistencias/>

¿Qué es el aprendizaje no supervisado? | IBM. (s. f.). <https://www.ibm.com/es-es/topics/unsupervised-learning>

¿Qué es el aprendizaje supervisado? | IBM. (s. f.). <https://www.ibm.com/mx-es/topics/supervised-learning>

¿Qué son los mercados financieros y cuáles son sus tipos? (s. f.). SNHU. <https://es.snhu.edu/noticias/definicion-y-tipos-de-mercados-financieros>

Roldán, P. N. (2023). Bolsa de valores. Economipedia. https://economipedia.com/definiciones/bolsa-de-valores.html?nab=1&utm_referrer=https%3A%2F%2Fwww.google.com%2F

RPubs - Distributional Regression Forest: Random Forest probabilístico. (s. f.). https://rpubs.com/Joaquin_AR/587903

Rübke, M. (2022, 24 septiembre). 40% de las empresas en LATAM utilizan Inteligencia Artificial. MADBOXPC. <https://www.madboxpc.com/40-de-las-empresas-en-latam-utilizan-inteligencia-artificial/>

Santander. (2022). ¿Qué son los índices bursátiles y para qué sirven? www.santander.com. <https://www.santander.com/es/stories/que-son-los-indices-bursatiles-y-para-que-sirven>

Santander, B. (s. f.-a). Ibex 35. Banco Santander. <https://www.bancosantander.es/glosario/ibex-35>

Santander, B. (s. f.-b). RSI o Índice de Fuerza Relativa. Banco Santander. <https://www.bancosantander.es/glosario/rsi-o-indice-de-fuerza-relativa>

septiembre 2022 – Grupo de Investigación y Desarrollo en Innovación Educativa. (2022, 15 septiembre). <https://gridie.exactas.unlpam.edu.ar/?m=202209>

Sergio, B. B. (s. f.). ANÁLISIS DE SERIES DE TIEMPO PARA LA PREDICCIÓN DE LOS PRECIOS DE LA ENERGÍA EN LA BOLSA DE COLOMBIA. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-47722008000100007

Series Temporales. (s. f.). <http://www5.uva.es/estadmed/datos/series/series1.htm>

Unir, V. (2023, 1 febrero). ¿Qué son las redes neuronales? Concepto y usos principales. UNIR. <https://www.unir.net/ingenieria/revista/redes-neuronales-artificiales/>

What is a Neural Network? (s. f.). TIBCO Software. <https://www.tibco.com/reference-center/what-is-a-neural-network>

What is a Random Forest? (s. f.). TIBCO Software. <https://www.tibco.com/reference-center/what-is-a-random-forest>

Wikipedia contributors. (2023). Machine learning. Wikipedia. https://en.wikipedia.org/wiki/Machine_learning

ANNEX I. SCRIPTS D'RSTUDIO

```
``{r}
setwd("C:/Users/Raquel/Documents/UNI/4rt curs/2n quadrimestre/TFG/R")
```

``{r}
library(quantmod)
library(dplyr)
library(tidyr)
library(ggplot2)
library(caret)
```

# Importació de les dades
``{r}
getSymbols("^IBEX", from = "2000-01-01", to = "2022-12-31")
chartSeries(IBEX, theme="white", TA=NULL)

# Creem el data frame
ibex35 <- as.data.frame(IBEX)
```

Tractament de les dades
``{r}
Convertim el nom de les files (la data) en una variable
ibex35$Date <- rownames(ibex35)
rownames(ibex35) <- NULL
str(ibex35)

Li posem format date
ibex35$Date <- as.Date(ibex35$Date)

Simplifiquem el nom de les variables
ibex35$Open <- ibex35$IBEX.Open
ibex35$High <- ibex35$IBEX.High
ibex35$Low <- ibex35$IBEX.Low
ibex35$Close <- ibex35$IBEX.Close
ibex35$Volume <- ibex35$IBEX.Volume
ibex35$Adjusted <- ibex35$IBEX.Adjusted

ibex35$IBEX.Open <- ibex35$IBEX.High <- ibex35$IBEX.Low <- ibex35$IBEX.Close <-
ibex35$IBEX.Volume <- ibex35$IBEX.Adjusted <- NULL

str(ibex35)
```
```

```
# Anàlisi descriptiva inicial
```

```
## Univariant
```

```
``{r}
```

```
columnes <- colnames(ibex35)[2:7]
```

```
# Summary
```

```
par(ask=FALSE)
```

```
for(k in columnes){
```

```
cat("\nVariable:", colnames(ibex35[k]), fill=TRUE)
```

```
print(summary(ibex35[,k]))
```

```
}
```

```
...
```

```
``{r}
```

```
ggplot(ibex35, aes(x = Date)) +
```

```
  geom_line(aes(y = Open, color = "Open")) +
```

```
  geom_line(aes(y = Close, color = "Close")) +
```

```
  geom_line(aes(y = High, color = "High")) +
```

```
  geom_line(aes(y = Low, color = "Low")) +
```

```
  # geom_line(aes(y = Volume, color = "Volume")) +
```

```
  geom_line(aes(y = Adjusted, color = "Adjusted")) +
```

```
  labs(x = "Dates", y = "Punts d'índex", color = "Variable") +
```

```
  scale_color_manual(values = c("Open" = "blue", "Close" = "red", "High" = "green", "Low" =  
"purple", "Adjusted" = "pink")) +
```

```
  ggtitle("Gràfic de les variables Open, Close, High, Low i Adjusted") + theme(plot.title =  
element_text(hjust = 0.5))
```

```
ggplot(ibex35, aes(x = Date)) +
```

```
  geom_line(aes(y = Volume, color = "Volume")) +
```

```
  labs(x = "Dates", y = "Volum", color = "Variable") +
```

```
  scale_color_manual(values = c("Volume" = "LightSkyBlue")) +
```

```
  ggtitle("Gràfic de la variable Volume") + theme(plot.title = element_text(hjust = 0.5))
```

```
...
```

```
Zoom abans de 2010
```

```
``{r}
```

```
ibex35_filtered <- ibex35 %>%
```

```
  filter(Date < as.Date("2010-01-01"))
```

```
ggplot(ibex35_filtered, aes(x = Date)) +
```

```
  geom_line(aes(y = Volume, color = "Volume")) +
```

```
  labs(x = "Dates", y = "Volum", color = "Variable") +
```

```
  scale_color_manual(values = c("Volume" = "LightSkyBlue")) +
```

```
  ggtitle("Gràfic de la variable Volume fins el 2010") + theme(plot.title = element_text(hjust =  
0.5))
```

```
...
```

Zoom entre 2002 i 2010

```
``{r}
ibex35_filtered2 <- ibex35 %>%
  filter(Date > as.Date("2002-01-01") & Date < as.Date("2010-01-01"))

ggplot(ibex35_filtered2, aes(x = Date)) +
  geom_line(aes(y = Volume, color = "Volume")) +
  labs(x = "Dates", y = "Volum", color = "Variable") +
  scale_color_manual(values = c("Volume" = "LightSkyBlue")) +
  ggtitle("Gràfic de la variable Volume entre el 2002 i el 2010") + theme(plot.title =
element_text(hjust = 0.5))
``
```

Bivariant

```
``{r}
# Correlació
library(corrplot)
library(Hmisc)
num <- ibex35[2:7]
correlacio <- round(cor(num, use = "complete.obs"),2)
corrplot(correlacio ,method = "number",type = "upper")
``
```

Preprocessing

```
## Tractament de missings
``{r}
# Observacions missings
(observacions_na <- ibex35$Date[is.na(ibex35$Open)])
```

```
# Mirem les dates que hi ha
library("lubridate")
df_missings <- data.frame(
  Any = year(observacions_na),
  Dates = format(observacions_na, "%d-%m-%Y")
)
```

```
# Tabla de recuento de días por mes
(tabla <- df_missings %>%
  group_by(Any) %>%
  summarise(Dates = paste(Dates, collapse = ", ")))
```

```
# Imputació missings mitjançant interpolació
ibex35interpolat <- na.approx(ibex35[2:7])
summary(ibex35interpolat)
ibex35new <- as.data.frame(ibex35interpolat)
```

```

Date <- ibex35$Date
ibex35new <- cbind(Date, ibex35new)
...

# Tractament d'outliers
``{r}
# Boxplot
par(mfrow = c(1,3))
for(k in columnes){
  grafic <- ggplot(ibex35new, aes(y = ibex35new[, k]))
  grafic <- grafic + geom_boxplot(color = "blue", fill = "lightblue")
  grafic <- grafic + labs(title = paste("Boxplot de", k),
    y = "Punts d'índex",
    x = k) + theme(plot.title = element_text(hjust = 0.5))
  print(grafic)

  # boxplot(ibex35[,k], main=paste("Boxplot de", colnames(ibex35[k])))
}
...

# Estadístics
## Medias mòviles centradas
``{r}
library("zoo")
mmc <- ibex35new %>%
  select(Date, Adjusted) %>%
  mutate(Ajustat_ma1_30 = rollmean(Adjusted, k = 30, fill = NA),
    Ajustat_ma2_182 = rollmean(Adjusted, k = 182, fill = NA),
    Ajustat_ma3_365 = rollmean(Adjusted, k = 365, fill = NA),
    Ajustat_ma4_550 = rollmean(Adjusted, k = 550, fill = NA),
    Ajustat_ma5_1000 = rollmean(Adjusted, k = 1000, fill = NA))
mmc %>%
  gather(Rangs, value, Adjusted:Ajustat_ma5_1000) %>%
  ggplot(aes(Date, value, color = Rangs)) + geom_line(size=0.8) + labs(title = paste("Mitjana
mòbil simple de la variable Adjusted"), y = "Punts d'índex", x = "Dates") + theme(plot.title =
element_text(hjust = 0.5))
...

``{r}
mmc %>%
  gather(metric, value, Adjusted:Ajustat_ma5_1000) %>%
  group_by(metric) %>%
  summarise(MSE = mean((mmc$Adjusted - value)^2, na.rm = TRUE),
    RMSE = sqrt(mean((mmc$Adjusted - value)^2, na.rm = TRUE)),
    MAE = mean(abs(mmc$Adjusted - value), na.rm = TRUE),
    MAPE = mean(abs((mmc$Adjusted - value)/mmc$Adjusted), na.rm = TRUE))
...

```

```
## Medias moviles exponenciales
```

```
``{r}
library("TTR")
ema <- ibex35new %>%
  select(Date, Adjusted) %>%
  mutate(Ajustat_ema1_30 = EMA(Adjusted, n = 30, fill = NA),
         Ajustat_ema2_182 = EMA(Adjusted, n = 182, fill = NA),
         Ajustat_ema3_365 = EMA(Adjusted, n = 365, fill = NA),
         Ajustat_ema4_550 = EMA(Adjusted, n = 550, fill = NA),
         Ajustat_ema5_1000 = EMA(Adjusted, n = 1000, fill = NA))
ema %>%
  gather(Rangs, value, Adjusted:Ajustat_ema5_1000) %>%
  ggplot(aes(Date, value, color = Rangs)) + geom_line(size=0.8) + labs(title = paste("Mitjana
mòbil exponencial de la variable Adjusted"), y = "Punts d'índex", x = "Dates") + theme(plot.title
= element_text(hjust = 0.5))
``
```

```
``{r}
ema %>%
  gather(metric, value, Adjusted:Ajustat_ema5_1000) %>%
  group_by(metric) %>%
  summarise(MSE = mean((ema$Adjusted - value)^2, na.rm = TRUE),
           RMSE = sqrt(mean((ema$Adjusted - value)^2, na.rm = TRUE)),
           MAE = mean(abs(ema$Adjusted - value), na.rm = TRUE),
           MAPE = mean(abs((ema$Adjusted - value)/ema$Adjusted), na.rm = TRUE))
``
```

```
## Medias moviles ponderadas
```

```
``{r}
wma <- ibex35new %>%
  select(Date, Adjusted) %>%
  mutate(Ajustat_wma1_30 = WMA(Adjusted, n = 30, fill = NA),
         Ajustat_wma2_182 = WMA(Adjusted, n = 182, fill = NA),
         Ajustat_wma3_365 = WMA(Adjusted, n = 365, fill = NA),
         Ajustat_wma4_550 = WMA(Adjusted, n = 550, fill = NA),
         Ajustat_wma5_1000 = WMA(Adjusted, n = 1000, fill = NA))
wma %>%
  gather(Rangs, value, Adjusted:Ajustat_wma5_1000) %>%
  ggplot(aes(Date, value, color = Rangs)) + geom_line(size=0.8) + labs(title = paste("Mitjana
mòbil ponderada de la variable Adjusted"), y = "Punts d'índex", x = "Dates") + theme(plot.title
= element_text(hjust = 0.5))
``
```

```
``{r}
wma %>%
  gather(metric, value, Adjusted:Ajustat_wma5_1000) %>%
```

```

group_by(metric) %>%
  summarise(MSE = mean((wma$Adjusted - value)^2, na.rm = TRUE),
            RMSE = sqrt(mean((wma$Adjusted - value)^2, na.rm = TRUE)),
            MAE = mean(abs(wma$Adjusted - value), na.rm = TRUE),
            MAPE = mean(abs((wma$Adjusted - value)/wma$Adjusted), na.rm = TRUE))
...

## Comparació
```{r}
compMA <- data.frame(Date = mmc$Date, Ajustat = mmc$Adjusted, Ajustat_SMA =
mmc$Ajustat_ma3_365, Ajustat_EMA = ema$Ajustat_ema3_365, Ajustat_WMA =
wma$Ajustat_wma3_365)

compMA %>%
 gather(Mitjanes, value, Ajustat:Ajustat_WMA) %>%
 ggplot(aes(Date, value, color = Mitjanes)) + geom_line(size=0.8) + labs(title =
paste("Comparació mitjanes mòbils de la variable Adjusted"), y = "Punts d'índex", x = "Dates")
+ theme(plot.title = element_text(hjust = 0.5))
...

```{r}
compMA %>%
  gather(metric, value, Ajustat:Ajustat_WMA) %>%
  group_by(metric) %>%
  summarise(MSE = mean((compMA$Ajustat - value)^2, na.rm = TRUE),
            RMSE = sqrt(mean((compMA$Ajustat - value)^2, na.rm = TRUE)),
            MAE = mean(abs(compMA$Ajustat - value), na.rm = TRUE),
            MAPE = mean(abs((compMA$Ajustat - value)/compMA$Ajustat), na.rm = TRUE))
...

## RSI
```{r}
rsiibex <- ibex35new %>%
 select(Date, Adjusted) %>%
 mutate(Ajustat_rsi1_30 = RSI(Adjusted, n = 30, fill = NA),
 Ajustat_rsi2_182 = RSI(Adjusted, n = 182, fill = NA),
 Ajustat_rsi3_365 = RSI(Adjusted, n = 365, fill = NA),
 Ajustat_rsi4_550 = RSI(Adjusted, n = 550, fill = NA),
 Ajustat_rsi5_1000 = RSI(Adjusted, n = 1000, fill = NA))
rsiibex %>%
 gather(Rangs, value, Ajustat_rsi1_30:Ajustat_rsi5_1000) %>%
 ggplot(aes(Date, value, color = Rangs)) + geom_line(size=0.8) + labs(title = paste("Índex de
força relativa (RSI)", y = "Valor", x = "Dates") + theme(plot.title = element_text(hjust = 0.5))
...

```



```

Xarxes neuronals

Preparació de les dades
```{r}
ibexajustat <- as.data.frame(cbind(ibex35new$Date, ibex35new$Adjusted))
str(ibexajustat)
ibexajustat$Date <- ibexajustat$V1
ibexajustat$Adjusted <- ibexajustat$V2
ibexajustat$Date <- as.Date(ibexajustat$Date)
ibexajustat$V1 <- ibexajustat$V2 <- NULL
str(ibexajustat)
```

Rang predicció
```{r}
rangdata <- seq(as.Date("2023-01-01"), as.Date("2023-01-31"), by = "day")
Adjusted <- as.numeric(NA)
prediccio <- as.data.frame(cbind(rangdata, Adjusted))
prediccio$Date <- as.Date(prediccio$rangdata)
prediccio$rangdata <- NULL
xarxesibex <- rbind(ibexajustat, prediccio)
```

Separar la data en dia, mes, any
```{r}
xarxesibex$Date_dup <- xarxesibex$Date
xarxesibex <- xarxesibex %>% separate(Date, c("Any", "Mes", "Dia"))
str(xarxesibex)
xarxesibex$Any <- as.numeric(xarxesibex$Any)
xarxesibex$Mes <- as.numeric(xarxesibex$Mes)
xarxesibex$Dia <- as.numeric(xarxesibex$Dia)
str(xarxesibex)
```

Escalat de dades (any, mes i dia)
```{r}
set.seed(20255244)
xarxesibexsc <- as.data.frame(cbind(xarxesibex$Adjusted, xarxesibex$Date_dup,
scale(xarxesibex[, c(1:3)])))
names(xarxesibexsc)[1] <- "Adjusted"
names(xarxesibexsc)[2] <- "Date"
xarxesibexsc$Date <- as.Date(xarxesibexsc$Date)
```

Separació de les dades
```{r}
set.seed(20255244)

```

```

train <- createDataPartition(na.omit(subset(xarxesibex, xarxesibex$Date_dup < "2023-01-01"))$Adjusted, p = 0.7, list = F)
test <- rbind(xarxesibex[-train,], subset(xarxesibex, xarxesibex$Date_dup >= "2023-01-01"))
testsc <- as.data.frame(cbind(test$Adjusted, test$Date_dup, scale(test[, c(1:3)])))
names(testsc)[1] <- "Adjusted"
names(testsc)[2] <- "Date"
testsc$Date <- as.Date(testsc$Date)
...

## Modelatge
``{r}
library(neuralnet)
library(NeuralNetTools)
...

``{r}
mod <- neuralnet(formula = Adjusted ~ Any + Mes + Dia, data = xarxesibexsc[train,], hidden = 2, threshold = 0.01, stepmax = 1e+08, rep = 1, linear.output = TRUE)

plotnet(mod)
garson(mod)
olden(mod)

# Garson
rel_imp <- garson(mod, bar_plot = FALSE)$rel_imp
cols <- colorRampPalette(c("lightgreen", "darkgreen"))(3)[rank(rel_imp)]
plotnet(mod, circle_col = list(cols, "lightblue"))
title("Xarxa neuronal")

# Olden
rel_imp <- olden(mod, bar_plot = FALSE)$importance
cols <- colorRampPalette(c("lightgreen", "darkgreen"))(3)[rank(rel_imp)]
plotnet(mod, circle_col = list(cols, "lightblue"))
...

## Resultats
``{r}
pred <- compute(mod, testsc)
xarxesresult <- cbind(pred$net.result, testsc)

ggplot() + geom_line(data = xarxesresult, aes(x = Date, y = Adjusted), color = "blue") +
geom_line(data = xarxesresult, aes(x = Date, y = `pred$net.result`), color = "#FF99FF", size=1)
+ labs(title = paste("Predicció de la variable Adjusted amb Xarxes Neuronals"), y = "Punts d'índex", x = "Dates") + theme(plot.title = element_text(hjust = 0.5))
...

```

```
## Comparació amb els resultats reals
```

```
``{r}  
library(DescTools)  
(msexn <-MSE(xarxesresult$Adjusted, xarxesresult$`pred$net.result`, na.rm = TRUE))  
(rmsexn <- RMSE(xarxesresult$Adjusted, xarxesresult$`pred$net.result`, na.rm = TRUE))  
(error_percentual_xn <- rmsexn/xarxesresult[xarxesresult$Date ==  
max(na.omit(xarxesresult)$Date),]$Adjusted)  
(maexn <- MAE(xarxesresult$Adjusted, xarxesresult$`pred$net.result`, na.rm = TRUE))  
(mapexn <- MAPE(xarxesresult$Adjusted, xarxesresult$`pred$net.result`, na.rm = TRUE))  
``
```

```
# Random forest
```

```
``{r}  
library(randomForest)  
``
```

```
## Modelatge
```

```
``{r}  
mod_rd <- randomForest(Adjusted ~ Any + Mes + Dia, data = xarxesibex[train,], type =  
"regression", ntree = 100)  
``
```

```
## Resultats
```

```
``{r}  
pred_rf <- predict(mod_rd, test)  
rfresult <- cbind(pred_rf, test)
```

```
ggplot() + geom_line(data = rfresult, aes(x = Date_dup, y =Adjusted), color ="blue") +  
geom_line(data = rfresult, aes(x = Date_dup, y = pred_rf), color ="#FF99FF", size=1) +  
labs(title = paste("Predicció de la variable Adjusted amb Random Forest"), y = "Punts d'índex",  
x = "Dates") + theme(plot.title = element_text(hjust = 0.5))  
``
```

```
## Comparació amb els resultats reals
```

```
``{r}  
(mserf <-MSE(rfresult$Adjusted, rfresult$pred_rf, na.rm = TRUE))  
(rmserf <- RMSE(rfresult$Adjusted, rfresult$pred_rf, na.rm = TRUE))  
(error_percentual_rf <- rmserf/rfresult[rfresult$Date_dup ==  
max(na.omit(rfresult)$Date_dup),]$Adjusted)  
(maerf <- MAE(rfresult$Adjusted, rfresult$pred_rf, na.rm = TRUE))  
(maperf <- MAPE(rfresult$Adjusted, rfresult$pred_rf, na.rm = TRUE))  
``
```

```

# Gradient Boosting
```{r}
library(gbm)
library(caret)
```

## Modelatge 100 arbres
```{r}
model_gbm = gbm(Adjusted ~ Any + Mes + Dia, data = xarxesibex[train,],
 distribution = "gaussian",
 cv.folds = 10,
 shrinkage = .01,
 n.minobsinnode = 10,
 n.trees = 100)

print(model_gbm)
summary(model_gbm)
```

## Resultats
```{r}
pred_gbm <- predict(model_gbm, test)
gbmresult <- cbind(pred_gbm, test)

ggplot() + geom_line(data = gbmresult, aes(x = Date_dup, y = Adjusted), color = "blue") +
 geom_line(data = gbmresult, aes(x = Date_dup, y = pred_gbm), color = "#FF99FF", size=1) +
 labs(title = paste("Predicció de la variable Adjusted amb Gradient Boosting"), y = "Punts
d'índex", x = "Dates") + theme(plot.title = element_text(hjust = 0.5))
```

## Comparació amb els resultats reals

```{r}
Errors
(msegbm <- MSE(gbmresult$Adjusted, gbmresult$pred_gbm, na.rm = TRUE))
(rmsegbm <- RMSE(gbmresult$Adjusted, gbmresult$pred_gbm, na.rm = TRUE))
(error_percentual_gbm <- rmsegbm/gbmresult[gbmresult$Date_dup ==
max(na.omit(gbmresult)$Date_dup),]$Adjusted)
(maegbm <- MAE(gbmresult$Adjusted, gbmresult$pred_gbm, na.rm = TRUE))
(mapegbm <- MAPE(gbmresult$Adjusted, gbmresult$pred_gbm, na.rm = TRUE))
```

## Modelatge 500 arbres
```{r}
model_gbm = gbm(Adjusted ~ Any + Mes + Dia, data = xarxesibex[train,],
 distribution = "gaussian",
 cv.folds = 10,
 shrinkage = .01,

```

```

n.minobsinnode = 10,
n.trees = 500)

print(model_gbm)
summary(model_gbm)
...

Resultats
``{r}
pred_gbm <- predict(model_gbm, test)
gbmresult <- cbind(pred_gbm, test)

ggplot() + geom_line(data = gbmresult, aes(x = Date_dup, y = Adjusted), color = "blue") +
geom_line(data = gbmresult, aes(x = Date_dup, y = pred_gbm), color = "#FF99FF", size=1) +
labs(title = paste("Predicció de la variable Adjusted amb Gradient Boosting"), y = "Punts
d'índex", x = "Dates") + theme(plot.title = element_text(hjust = 0.5))
...

Comparació amb els resultats reals

``{r}
Errors
(msegbm <- MSE(gbmresult$Adjusted, gbmresult$pred_gbm, na.rm = TRUE))
(rmsegbm <- RMSE(gbmresult$Adjusted, gbmresult$pred_gbm, na.rm = TRUE))
(error_percentual_gbm <- rmsegbm/gbmresult[gbmresult$Date_dup ==
max(na.omit(gbmresult)$Date_dup),]$Adjusted)
(maegbm <- MAE(gbmresult$Adjusted, gbmresult$pred_gbm, na.rm = TRUE))
(mapegbm <- MAPE(gbmresult$Adjusted, gbmresult$pred_gbm, na.rm = TRUE))
...

```