
This is the **published version** of the bachelor thesis:

Lozano Amores, Aaron. *Reproducció i avaluació del mètode “Trading via Image Classification” per a la predicció de patrons en sèries temporals financeres..*

Treball de Final de Grau (Universitat Autònoma de Barcelona), 2026
(Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/334405>

under the terms of the  license.

Classificació de senyals tècnics de compra a l'S&P 500 mitjançant imatges de veles japoneses i aprenentatge automàtic

Aaron Lozano Amores

29 de juny de 2026

Resum– Aquest treball estudia la classificació supervisada de senyals tècnics de compra en actius de l'S&P 500 a partir de dades financeres històriques. Les sèries de preus OHLC es transformen en finestres temporals i posteriorment en imatges de veles japoneses de 30×30 píxels. Cada imatge s'etiqueta mitjançant regles basades en tres indicadors tècnics: RSI, MACD i Bandes de Bollinger. L'objectiu no és predir directament el comportament futur del mercat, sinó analitzar si els models poden reconèixer visualment patrons associats a aquests senyals. Per fer-ho, es comparen models clàssics d'aprenentatge automàtic, una xarxa neuronal convolucional i un sistema de votació per majoria. Els resultats mostren que el rendiment depèn de l'indicador utilitzat, que el MACD és el cas més difícil i que la CNN obté els millors resultats globals en termes d'F1-score. A més, s'analitzen els falsos positius i s'incorpora una avaluació borsària simplificada dels senyals generats.

Paraules clau– Classificació supervisada, aprenentatge automàtic, S&P 500, veles japoneses, indicadors tècnics, xarxes neuronals convolucionals.

Abstract– This work studies the supervised classification of technical buy signals in S&P 500 assets using historical financial data. OHLC price series are transformed into temporal windows and then converted into 30×30 pixel candlestick chart images. Each image is labelled using rules based on three technical indicators: RSI, MACD and Bollinger Bands. The aim is not to directly predict future market behaviour, but to analyse whether models can visually recognise patterns associated with these signals. To do so, several classical machine learning models, a convolutional neural network and a majority voting system are compared. The results show that performance depends on the technical indicator used, that MACD is the most difficult case, and that the CNN obtains the best overall results in terms of F1-score. In addition, false positives are analysed and a simplified financial evaluation of the generated signals is included.

Keywords– Supervised classification, machine learning, S&P 500, candlestick charts, technical indicators, convolutional neural networks.



1 INTRODUCCIÓ - CONTEXT DEL TREBALL

ELS mercats financers generen una gran quantitat de dades al llarg del temps. Aquestes dades es poden estudiar de moltes maneres, ja sigui amb va-

lors numèrics, amb indicadors tècnics o amb gràfics. En aquest treball s'ha volgut treballar especialment amb aquesta última idea, és a dir, amb la representació visual de les dades financeres.

Els gràfics de veles japoneses són molt utilitzats en l'anàlisi tècnica perquè resumeixen, en cada sessió, el preu d'obertura, el màxim, el mínim i el preu de tancament d'un actiu. Això permet veure de manera visual com ha evolucionat el preu durant un període concret. A partir d'aquesta representació, el projecte planteja si és possible fer servir models de classificació per reconèixer patrons associats a possibles senyals tècnics de compra.

- E-mail de contacte: 1636164@uab.cat
- Menció: Computació
- Treball tutoritzat per: Albert Romero Sanchez (Ciències de la computació i intel·ligència artificial)
- Curs 2025/26

La idea del treball no és predir directament què farà el mercat en el futur. El que es vol analitzar és si, a partir d'imatges de veles japoneses, els models poden distingir entre finestres temporals on apareix un senyal tècnic de compra i finestres on aquest senyal no apareix. Per tant, el problema es formula com una tasca de classificació supervisada basada en imatges, i no com una predicció directa de la rendibilitat futura d'un actiu.

Per construir aquestes imatges s'utilitzen dades històriques d'actius de l'S&P 500. Les dades originals contenen els valors OHLC, és a dir, obertura, màxim, mínim i tancament. A partir d'aquestes dades es calculen tres indicadors tècnics: RSI, MACD i Bandes de Bollinger [13, 14, 15]. Cada indicador defineix una regla pròpia per marcar si en una finestra temporal hi ha, o no, un possible senyal de compra.

Un cop generades les imatges i les etiquetes, es comparen diferents models de classificació. En concret, s'avaluen diversos models clàssics d'aprenentatge automàtic, una xarxa neuronal convolucional i un sistema de votació per majoria. També s'analitzen els falsos positius i s'afegeix una avaluació borsària simplificada dels senyals generats, per no limitar l'anàlisi només a les mètriques de classificació.

La resta de l'article s'organitza de la manera següent. Primer es presenta l'estat de l'art i els treballs relacionats. Després s'explica la metodologia seguida per obtenir les dades, preparar-les, generar les imatges i entrenar els models. A continuació es mostren els resultats obtinguts i la seva discussió. Finalment, es presenten les conclusions, les limitacions del treball i possibles línies futures.

2 ESTAT DE L'ART

L'anàlisi dels mercats financers amb tècniques computacionals s'ha treballat sovint a partir de sèries temporals de preus, indicadors tècnics i models d'aprenentatge automàtic. Indicadors com el RSI, el MACD o les Bandes de Bollinger permeten resumir diferents aspectes del comportament recent d'un actiu, com el moment del preu, els canvis de tendència o la relació del preu amb la seva volatilitat [13, 14, 15]. En aquest treball, aquests indicadors no s'utilitzen com una garantia de predicció del mercat, sinó com a criteri per definir etiquetes de compra o no compra de manera clara i reproducible.

Una línia de treball relacionada consisteix a transformar dades financeres en imatges. Cohen et al.[1] plantegen l'ús de representacions visuals generades a partir de dades de mercat per abordar el problema com una tasca de classificació d'imatges. Aquesta idea és rellevant perquè permet passar d'una representació purament numèrica a una representació visual, on els models poden intentar detectar patrons en la forma del moviment del preu.

També hi ha treballs centrats en l'ús de gràfics de veles japoneses com a entrada per a xarxes neuronals convolucionals. Chen i Tsai [2] estudien la codificació de patrons de veles com a imatges i avaluen la capacitat de la CNN per classificar aquest tipus de representacions. Aquest enfocament és proper al d'aquest TFG, ja que les veles japoneses resumeixen visualment els valors d'obertura, màxim, mínim i tancament de cada sessió.

Dins d'aquest context, aquest treball no presenta l'ús d'imatges financeres com una idea completament nova. La

seva aportació principal és construir i avaluar un flux propi aplicat a actius de l'S&P 500, amb etiquetes generades a partir de tres indicadors tècnics diferents: RSI, MACD i Bandes de Bollinger. A més, no s'avalua només una CNN, sinó que també es comparen models clàssics d'aprenentatge automàtic i un sistema de votació per majoria.

Així, el treball adapta idees ja presents en la literatura a un cas experimental concret, on es vol analitzar si els senyals tècnics de compra poden ser reconeguts a partir de la informació visual continguda en imatges de veles japoneses.

3 METODOLOGIA

La metodologia del treball s'ha estructurat com un procés per fases. La idea principal és partir de dades financeres diàries, preparar-les, calcular els senyals tècnics i acabar generant imatges que puguin ser utilitzades pels models de classificació.

El procés no comença directament amb les imatges. Primer es treballa amb dades tabulars de preus. Després, aquestes dades es netegen i s'ordenen. A continuació es calculen els indicadors tècnics i es defineixen les etiquetes de compra o no compra. Finalment, cada finestra temporal es converteix en una imatge de veles japoneses. D'aquesta manera, el problema passa de ser una sèrie temporal de preus a ser un problema de classificació d'imatges.

Aquest flux s'ha implementat de manera modular dins del Jupyter. Cada fase genera fitxers intermedis, de manera que es pot revisar què ha passat en cada pas sense haver de repetir tot el procés des del principi. Això també facilita detectar errors, comprovar la qualitat de les dades i mantenir una traçabilitat clara entre les dades originals i les imatges finals.

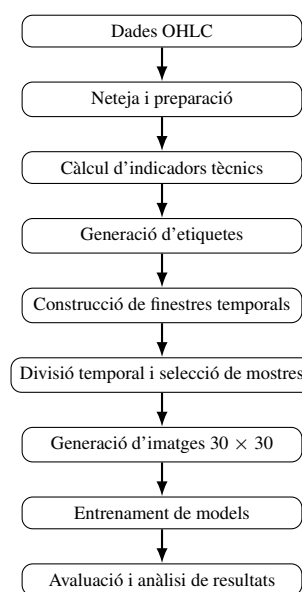


Fig. 1: Flux general del procés seguit per passar de dades financeres històriques a imatges etiquetades i avaluar els models de classificació.

3.1 Obtenció de les dades

Les dades utilitzades provenen d'actius de l'S&P 500. Per a cada actiu s'han descarregat preus històrics diaris mitjançant la llibreria *yfinance* [3]. El període considerat va de l'1 de gener de 2015 a l'1 de gener de 2026. Aquesta finestra temporal permet disposar de diversos anys de dades i inclou etapes de mercat diferents, cosa que ajuda a generar més varietat de mostres.

En l'experiment final s'ha treballat amb 250 tickers. Aquesta ampliació és important perquè alguns indicadors, especialment l'RSI, no generen senyals positius de manera gaire freqüent. Si s'utilitzen poques empreses, el nombre de finestres positives pot ser insuficient per entrenar i avaluar els models amb garanties. Per aquest motiu, es va ampliar el nombre d'actius respecte a les proves inicials del projecte.

De cada actiu s'han descarregat les variables habituals de mercat disponibles a Yahoo Finance, però per a la generació de les imatges s'han utilitzat principalment els valors OHLC: obertura, màxim, mínim i tancament. Aquests quatre valors són necessaris perquè cada vela japonesa representa exactament aquesta informació en una sessió concreta.

Un cop descarregades les dades, s'afegeix el ticker corresponent a cada registre i es construeix una taula única amb totes les observacions. Això permet treballar amb totes les empreses dins d'un mateix flux de processament, però mantenint sempre la separació per actiu quan es calculen indicadors, finestres i etiquetes.

3.2 Preparació de les dades

La preparació de les dades és una part important del treball, perquè qualsevol error en aquesta fase podria afectar totes les etapes posteriors. Abans de calcular indicadors o generar imatges, les dades s'han ordenat per ticker i data, respectant sempre l'ordre cronològic. Això és necessari perquè el projecte treballa amb sèries temporals i una alteració en l'ordre dels dies podria generar finestres incorrectes.

Després s'ha aplicat una revisió de qualitat sobre les dades OHLC. En primer lloc, s'han comprovat els valors nuls. Les files sense informació suficient en les variables d'obertura, màxim, mínim o tancament s'han eliminat, ja que no permeten construir correctament una vela japonesa. També s'han revisat possibles registres duplicats per evitar que una mateixa sessió aparegui més d'una vegada per al mateix actiu.

A continuació s'han comprovat incoherències bàsiques entre preus. Per exemple, el valor màxim d'una sessió no pot ser inferior al valor mínim. De la mateixa manera, els preus d'obertura i tancament han d'estar dins del rang format pel màxim i el mínim del dia. Aquestes comprovacions serveixen per eliminar registres que podrien generar imatges incorrectes o distorsionar el càlcul dels indicadors.

També s'ha mantingut una estructura de fitxers intermedis per separar les fases del procés. Primer es guarda una versió ordenada de les dades, després una versió neta sense valors OHLC no vàlids i, finalment, una versió preparada per calcular indicadors. Aquesta organització ajuda a revisar el procés pas a pas i evita barrejar dades originals amb dades ja transformades.

Després d'aquesta fase, el conjunt de dades queda prepa-

rat per calcular els indicadors tècnics. Per tant, la preparació no modifica l'objectiu del problema, sinó que assegura que les dades que entren a les fases següents siguin coherents, ordenades i utilitzables.

3.3 Càlcul dels indicadors i generació d'etiquetes

Després de preparar les dades OHLC, el següent pas consisteix a calcular els indicadors tècnics que s'utilitzen per definir les etiquetes del problema. En aquest treball s'han utilitzat tres indicadors: RSI, MACD i Bandes de Bollinger. Cada un d'ells genera una regla concreta per marcar si en un dia determinat apareix, o no, un possible senyal tècnic de compra [13, 14, 15].

Els indicadors s'han calculat de manera independent per a cada ticker. Això és important perquè cada actiu té la seva pròpia sèrie temporal. Per tant, no es barregen preus d'empreses diferents en el càlcul dels indicadors. Abans de calcular qualsevol senyal, les dades es mantenen agrupades per actiu i ordenades per data.

El primer indicador utilitzat és l'RSI. Aquest indicador mesura la força relativa dels moviments recents del preu i s'ha calculat amb el període habitual de 14 sessions. En aquest treball es considera que hi ha un possible senyal de compra quan l'RSI és inferior a 30. Aquest llindar s'associa habitualment a situacions de possible sobreventa.

El segon indicador és el MACD. Aquest indicador es calcula a partir de dues mitjanes mòbils exponencials, una de curt termini i una de llarg termini. En concret, s'ha utilitzat la configuració habitual de 12, 26 i 9 períodes. Es marca un possible senyal de compra quan la línia MACD creua a l'alça la seva línia de senyal. És a dir, quan el dia anterior el MACD estava per sota o igual que la línia de senyal, i el dia actual passa a estar per sobre.

El tercer indicador són les Bandes de Bollinger. Aquest indicador es calcula a partir d'una mitjana mòbil simple de 20 sessions i dues bandes situades al voltant d'aquesta mitjana. En aquest treball s'ha utilitzat la configuració habitual de dues desviacions estàndard. Es considera que hi ha un possible senyal de compra quan el preu de tancament és igual o inferior a la banda inferior.

A partir d'aquestes regles es generen tres variables binàries de senyal, una per a cada indicador. La classe 1 indica que es compleix la condició de compra definida per l'indicador corresponent, mentre que la classe 0 indica que aquesta condició no es compleix. D'aquesta manera, el problema no es construeix com una única classificació general, sinó com tres problemes de classificació separats: un per a l'RSI, un per al MACD i un per a les Bandes de Bollinger.

Taula 1: Regles utilitzades per generar les etiquetes positives segons cada indicador tècnic.

Indicador	Períodes	Condició de compra
RSI	14	$RSI < 30$
MACD	12, 26, 9	Encreuament alcista de la línia MACD sobre la línia de senyal
Bandes de Bollinger	20	$\text{Preu de tancament} \leq \text{banda inferior}$

3.4 Construcció de finestres temporals

Un cop calculades les etiquetes diàries, les sèries de preus es divideixen en finestres temporals. En aquest treball, una finestra és un conjunt consecutiu de sessions borsàries d'un mateix ticker. Aquesta finestra és la unitat que després es transforma en una imatge de veles japoneses.

El procés es fa de manera separada per a cada ticker i per a cada indicador. Primer es prenen les dades d'un actiu concret, ordenades per data, i en acabat es generen finestres per a l'RSI, el MACD i les Bandes de Bollinger de manera independent. Això permet mantenir separades les etiquetes de cada indicador i evitar barrejar criteris diferents dins d'un mateix conjunt de dades.

Cada finestra conté els valors OHLC de diversos dies consecutius. Si la finestra acaba en el dia t , l'etiqueta assignada a aquella mostra és la que correspon al dia t . Per tant, la imatge mostra el comportament del preu durant els dies anteriors i la classe indica si al final d'aquell període apareix, o no, el senyal tècnic de compra.

La mida de la finestra no és igual per a tots els indicadors. En aquest treball s'han utilitzat finestres de 20 sessions per a les Bandes de Bollinger, 26 sessions per al MACD i 27 sessions per a l'RSI. Aquest criteri segueix l'enfocament de Cohen et al. [1], on les imatges es retallen segons el nombre de sessions efectives associades a cada regla d'etiquetatge.

En el cas de les Bandes de Bollinger, la finestra de 20 sessions coincideix amb la mitjana mòbil utilitzada per calcular l'indicador. En el cas del MACD, la finestra de 26 sessions recull la mitjana exponencial llarga que forma part del càlcul. En el cas de l'RSI, l'indicador s'ha calculat amb 14 sessions, però la imatge s'ha construït amb 27 sessions seguint el criteri del treball de referència. Per tant, les 27 sessions no s'utilitzen per calcular l'RSI, sinó per definir el fragment de sèrie temporal que es converteix en imatge.

Aquesta decisió no modifica les regles d'etiquetatge. Una mostra d'RSI continua sent positiva quan l'RSI és inferior a 30, una mostra de MACD quan es produeix l'encreuament alcista i una mostra de Bandes de Bollinger quan el tancament és igual o inferior a la banda inferior. El que canvia entre indicadors és la quantitat d'informació temporal que es mostra a la imatge.

Per a cada finestra es guarda el ticker, l'indicador, la data inicial, la data final i l'etiqueta corresponent. Aquesta informació és necessària per generar les imatges, dividir posteriorment el conjunt de dades i revisar possibles errors o superposicions entre mostres.

3.5 Selecció de mostres i divisió del conjunt de dades

Un cop construïdes les finestres temporals, el següent pas és dividir-les en els subconjunts d'entrenament, validació i prova. Aquesta divisió no s'ha fet de manera aleatòria simple, sinó respectant l'ordre temporal de les dades. En un problema financer això és important, perquè no seria correcte entrenar un model amb informació posterior i avaluar-lo amb dades anteriors.

La divisió s'ha aplicat de manera separada per a cada indicador. Això vol dir que les finestres de l'RSI, del MACD i de les Bandes de Bollinger es tracten com tres conjunts diferents. D'aquesta manera, cada model s'entrena i s'ava-

lua només amb les mostres associades a un únic indicador, evitant barrejar etiquetes generades amb regles diferents.

Inicialment, les finestres de cada indicador s'ordenen per data final. A partir d'aquest ordre, les mostres més antigues es destinen al conjunt d'entrenament, les següents al conjunt de validació i les més recents al conjunt de proves. La proporció utilitzada és aproximadament del 70% per a entrenament, el 15% per a validació i el 15% per a prova.

Després d'aquesta divisió temporal, s'aplica una purga antisolapament entre subconjunts. Aquest pas és necessari perquè les finestres es construeixen amb dies consecutius i, per tant, dues finestres properes poden compartir una part important de les mateixes sessions. Si una finestra del conjunt de proves compartís dies amb una finestra del conjunt d'entrenament, l'avaluació podria quedar esbiaixada. Per reduir aquest risc, s'eliminen les finestres de validació i prova que se superposen amb finestres anteriors del mateix ticker i indicador.

Un altre aspecte important és el desequilibri inicial entre classes. En les dades originals, les finestres sense senyal de compra són molt més freqüents que les finestres amb senyal. Això és esperable, perquè els senyals tècnics de compra no apareixen cada dia. Si es mantingués aquest desequilibri, els models podrien aprendre sobretot a predir la classe majoritària, és a dir, la classe sense senyal.

Per aquest motiu, dins de cada indicador i subconjunt s'ha fet una selecció balancejada de mostres positives i negatives. En el conjunt d'entrenament s'han seleccionat el mateix nombre de finestres de classe 0 i de classe 1. El mateix criteri s'ha aplicat als conjunts de validació i prova. D'aquesta manera, cada model rep un conjunt equilibrat i es pot comparar millor el seu rendiment entre classes.

La selecció s'ha fet amb una llavor fixa per garantir que el procés sigui reproducible. Això vol dir que, si s'executa el Jupyter amb la mateixa configuració, s'obté la mateixa selecció de finestres. A més, en aquesta fase es conserva l'informació de cada mostra, com el ticker, l'indicador, la data inicial, la data final i l'etiqueta. Aquesta informació és útil per generar posteriorment les imatges i també per revisar possibles errors o falsos positius.

La distribució final de mostres obtinguda després d'aquest procés es presenta a l'apartat de resultats. En aquesta secció només s'ha descrit el criteri utilitzat per construir el conjunt de dades, mentre que la mida final de cada subconjunt es comenta més endavant.

3.6 Generació de les imatges de veles japoneses

Després de seleccionar les finestres temporals, aquestes es converteixen en imatges de veles japoneses. Aquest pas és necessari perquè els models no treballin directament amb la taula OHLC, sinó amb una representació visual del moviment del preu.

Per generar cada imatge s'utilitzen els valors d'obertura, màxim, mínim i tancament de la finestra corresponent. Cada dia es representa amb una vela: el cos mostra la diferència entre obertura i tancament, i la línia vertical mostra el rang entre el màxim i el mínim. Així, una finestra temporal queda resumida en una imatge amb l'evolució recent del preu.

Els gràfics s'han generat amb la llibreria *mplfinance* [11]. En les imatges no s'han inclòs eixos, títols, dates, volum ni cap text identificatiu. D'aquesta manera, el model només rep la forma de les veles i no pot aprendre informació externa com el ticker o la data. Aquest criteri és similar al de Cohen et al. [1], on les imatges també es construeixen sense etiquetes textuais.

Un cop generat el gràfic, la imatge es guarda en format RGB i es redimensiona a 30×30 píxels. Aquesta mida s'ha triat seguint Cohen et al. [1], que observen que el rendiment dels classificadors s'estabilitza al voltant d'aquesta resolució. A més, una mida reduïda permet treballar amb menys càrrega computacional. En el cas dels models clàssics, cada imatge RGB de 30×30 píxels es converteix després en un vector de 2.700 valors.



Fig. 2: Exemple d'imatge de veles japoneses generada a partir d'una finestra temporal de l'empresa Apple Inc. (AAPL) associada a les Bandes de Bollinger.

Per facilitar la revisió visual dels exemples, les imatges incloses en el document es mostren ampliades respecte a la seva mida original. D'aquesta manera es poden apreciar millor les veles i el patró general representat, tot i que els models treballen amb les imatges en la resolució definida de 30×30 píxels.

3.7 Models i configuració experimental

Un cop generades les imatges, s'han entrenat diferents models de classificació per comparar diversos enfocaments sobre el mateix problema. L'entrenament s'ha fet sempre de manera separada per indicador. Per tant, hi ha un conjunt d'imatges per a l'RSI, un per al MACD i un per a les Bandes de Bollinger. Això evita barrejar etiquetes que provenen de regles diferents.

Primer s'han provat models clàssics d'aprenentatge automàtic amb *scikit-learn* [7, 8]. Per poder utilitzar aquests models, cada imatge RGB de 30×30 píxels s'ha convertit en un vector numèric. Com que cada imatge té tres canals de color, cada mostra queda representada per 2.700 valors normalitzats.

Els models clàssics provats inclouen Logistic Regression, Gaussian Naive Bayes, LDA, QDA, Gaussian Process, k-NN, SVM lineal, SVM amb nucli RBF, SVM amb nucli

RBF i probabilitat, MLP, Decision Tree, Random Forest, Extra Trees, AdaBoost, Bagging i Gradient Boosting. No tots aquests models tenen la mateixa complexitat ni aprenen de la mateixa manera. Alguns treballen amb fronteres de decisió lineals, altres amb arbres, altres amb distàncies entre mostres i altres amb combinacions de classificadors.

Per a cada model s'han provat diverses configuracions d'hiperparàmetres. Cada configuració s'ha entrenat amb el conjunt d'entrenament i s'ha comparat amb el conjunt de validació. A partir d'aquesta validació s'ha seleccionat la millor configuració de cada model. Després, el conjunt de proves només s'ha utilitzat per obtenir l'avaluació final, no per ajustar el model.

També s'ha entrenat una xarxa neuronal convolucional amb PyTorch [9, 10]. En aquest cas, la imatge no s'aplanava abans d'entrar al model, sinó que es manté amb la seva estructura original de $3 \times 30 \times 30$. La CNN utilitza capes convolucionals per extreure patrons locals de la imatge, capes de pooling per reduir la mida de les representacions i una part final densa per fer la classificació binària.

Durant l'entrenament de la CNN s'han provat diferents configuracions, modificant paràmetres com la taxa d'aprenentatge, la mida del lot, el dropout, el nombre de canals i l'ús de batch normalization. La selecció de la configuració final s'ha fet a partir del conjunt de validació, de la mateixa manera que amb els models clàssics.

Finalment, s'ha implementat un sistema de votació per majoria, anomenat Hard Voting. En aquest cas, diversos models fan una predicció sobre la mateixa imatge i la classe final es decideix segons el vot majoritari. Aquest sistema permet comprovar si combinar diferents classificadors dona un resultat més estable que utilitzar un únic model.

Per completar la comparació, també s'ha tingut en compte la mida i la complexitat dels models principals. En el cas de la CNN, aquesta complexitat es pot expressar mitjançant el nombre de paràmetres entrenables. La xarxa utilitzada en l'avaluació final té 101.569 paràmetres entrenables, amb una arquitectura formada per tres blocs convolucionals de 32, 64 i 128 canals, i una capa densa de 64 unitats.

En els models clàssics, la complexitat s'ha valorat a partir dels hiperparàmetres principals. En el cas del Random Forest, els models seleccionats utilitzen 100 arbres. Per a les Bandes de Bollinger s'ha utilitzat una profunditat màxima de 20, mentre que per al MACD no s'ha limitat la profunditat màxima. En el cas de l'RBF SVM, utilitzat com a millor model per a l'RSI, la configuració principal és $C=10$ i $\gamma=\text{scale}$. Finalment, el Hard Voting no afegeix nous paràmetres entrenables, sinó que combina les prediccions de 17 models diferents.

En tots els casos s'ha mantingut la mateixa separació entre entrenament, validació i prova. Això és important perquè la comparació entre models sigui coherent. Si cada model s'avalués amb dades diferents, els resultats no serien comparables.

3.8 Mètriques d'avaluació

Per comparar els models s'han utilitzat diverses mètriques de classificació. No s'ha tingut en compte només l'accuracy, perquè aquesta mètrica pot donar una visió massa general del resultat. En aquest treball interessa saber també si

el model detecta correctament els senyals positius i si genera molts falsos positius.

La primera mètrica feta servir és l'accuracy, que indica el percentatge total de prediccions correctes. Tot i això, aquesta mètrica no diferencia entre encerts de la classe 0 i encerts de la classe 1. Per aquest motiu també s'han calculat la precisió, el recall i l'F1-score [16, 17].

La precisió mesura, de totes les mostres que el model ha predit com a senyal de compra, quantes realment ho eren. Aquesta mètrica és important perquè un valor baix indicaria que el model genera moltes alertes falses. El recall, en canvi, mesura quantes dels senyals positius reals és capaç de detectar. Finalment, l'F1-score combina precisió i recall en un únic valor, i és útil quan es vol resumir l'equilibri entre totes dues.

També s'ha utilitzat la balanced accuracy [18]. Aquesta mètrica calcula el rendiment tenint en compte les dues classes de manera més equilibrada. Encara que el conjunt final fet servir per entrenar i avaluar els models està balancejat, aquesta mètrica ajuda a comprovar que el model no funciona bé només en una de les dues classes.

Una altra mètrica considerada és el Matthews Correlation Coefficient (MCC) [19]. Aquesta mètrica resumeix la qualitat de la classificació tenint en compte els quatre valors de la matriu de confusió: veritables positius, veritables negatius, falsos positius i falsos negatius. El seu valor pot anar de -1 a 1, on 1 indica una classificació perfecta, 0 un comportament proper a l'atzar i -1 una classificació completament incorrecta.

A més de les mètriques agregades, s'han analitzat les matrius de confusió. Aquestes matrius permeten veure de manera més clara quins errors comet cada model. En aquest treball s'ha prestat una atenció especial als falsos positius, és a dir, als casos en què el model prediu un senyal de compra, però la regla tècnica no la marca com a tal. Aquest tipus d'error és rellevant perquè, en un context financer, podria interpretar-se com una entrada al mercat que no estava justificada per l'indicador.

Finalment, per complementar l'avaluació de classificació, s'ha fet una prova borsària simplificada amb les prediccions de la CNN. Aquesta prova no substitueix un back-testing complet, però permet observar quin comportament tindrien els senyals generats pel model sota una regla senzilla d'entrada i sortida. En aquesta avaluació s'utilitzen dades no balancejades, més properes a una situació real, on els senyals de compra són menys freqüents que els dies sense senyal.

La regla feta servir consisteix a entrar a l'obertura de l'endemà al senyal predit i sortir al tancament cinc sessions després. A partir d'aquestes operacions s'han calculat mètriques habituals en sistemes de trading, com el retorn total, el CAGR, el Sharpe Ratio, el Sortino Ratio i el Maximum Drawdown [21, 22, 23]. Aquestes mètriques s'utilitzen només com una anàlisi complementària, ja que no es consideren comissions, lliscament de preus ni una gestió completa del capital.

4 RESULTATS I DISCUSSIÓ

En aquesta secció es presenten els resultats obtinguts després d'aplicar el procés descrit anteriorment. Primer es mostra la distribució final del conjunt de dades generat.

Després es comparen els resultats dels models per a cada indicador tècnic i, finalment, s'analitzen els errors més rellevants, especialment els falsos positius. També s'inclou una avaluació borsària simplificada per observar el comportament dels senyals generats pel model en un escenari més proper a dades no balancejades.

4.1 Conjunt de dades final

Després de construir les finestres temporals, aplicar la divisió temporal i seleccionar les mostres balancejades, s'ha obtingut un conjunt final de 23.400 imatges. Aquest total correspon a la suma dels tres indicadors tècnics utilitzats: RSI, MACD i Bandes de Bollinger.

Tot i això, aquest nombre global s'ha d'interpretar amb cura. Els models no s'han entrenat amb totes les imatges barrejades, sinó de manera separada per indicador. Per tant, cada indicador disposa del seu propi conjunt d'entrenament, validació i prova. Aquesta separació és important perquè cada etiqueta prové d'una regla tècnica diferent.

Per a cada indicador s'han generat 7.800 imatges. La distribució final és la mateixa en els tres casos: 5.600 imatges per a entrenament, 1.100 per a validació i 1.100 per a prova. A més, cada subconjunt està balancejat entre la classe 0 i la classe 1. Això permet comparar millor el comportament dels models en les dues classes.

Taula 2: Distribució final de mostres per indicador, subconjunt i classe.

Indicador	Entrenament	Validació	Prova
RSI	2.800 / 2.800	550 / 550	550 / 550
MACD	2.800 / 2.800	550 / 550	550 / 550
Bandes de Bollinger	2.800 / 2.800	550 / 550	550 / 550

Cada cel·la indica mostres de classe 0 / classe 1.

En cada indicador, el conjunt d'entrenament conté 2.800 mostres de classe 0 i 2.800 mostres de classe 1. Els conjunts de validació i prova contenen 550 mostres de cada classe. D'aquesta manera, l'avaluació final no queda condicionada per una classe majoritària dins del conjunt balancejat.

Aquesta distribució també facilita la comparació entre indicadors. Com que l'RSI, el MACD i les Bandes de Bollinger tenen el mateix nombre de mostres en entrenament, validació i prova, les diferències observades en els resultats es poden atribuir més fàcilment a la dificultat de cada regla tècnica i no a una diferència en la mida del conjunt de dades.

4.2 Comparació general dels models

La Taula 3 resumeix els millors resultats obtinguts per a cada indicador tècnic. Per no fer la taula massa extensa, es mostren els models més rellevants de cada cas: el millor model individual i el sistema de votació per majoria. En alguns indicadors també s'inclou un model clàssic destacat, per poder comparar millor el resultat de la CNN amb altres enfocaments.

Els resultats mostren que la CNN obté el millor F1-score en els tres indicadors. Aquest resultat és coherent amb el tipus de dades utilitzades, ja que la CNN treballa directament amb l'estructura espacial de la imatge. A diferència dels models clàssics, no necessita convertir la imatge en un

Taula 3: Principals resultats obtinguts en el conjunt de prova.

Indicador	Model	Accuracy	F1	MCC
Bollinger	CNN	0,9809	0,9809	0,9618
Bollinger	Random Forest	0,9636	0,9642	0,9278
Bollinger	Hard Voting	0,9673	0,9677	0,9349
MACD	CNN	0,7900	0,8051	0,5871
MACD	Random Forest	0,7955	0,7993	0,5913
MACD	Hard Voting	0,7964	0,7975	0,5928
RSI	CNN	0,9755	0,9757	0,9510
RSI	RBF SVM	0,9700	0,9702	0,9401
RSI	Hard Voting	0,9682	0,9688	0,9372

vector pla, sinó que pot aprendre patrons locals entre píxels i formes de les veles.

Tot i això, els models clàssics també obtenen resultats competitiu. En el cas de les Bandes de Bollinger, Random Forest arriba a un F1-score de 0,9642, molt proper al resultat de la CNN. En el cas de l'RSI, el model RBF SVM també aconsegueix un rendiment alt, amb un F1-score de 0,9702. Això indica que, encara que la representació sigui visual, alguns models clàssics poden separar força bé les classes quan les imatges són petites i estan normalitzades.

El cas del MACD és diferent. Els resultats són clarament més baixos que en RSI i Bandes de Bollinger. La CNN obté un F1-score de 0,8051, mentre que Hard Voting arriba a 0,7975 i Random Forest a 0,7993. Per tant, cap model aconsegueix un rendiment tan elevat com en els altres dos indicadors. Això suggereix que la regla del MACD és més difícil de recuperar només a partir de la forma de les veles.

Aquesta diferència té sentit si es té en compte com es construeixen les etiquetes. Les Bandes de Bollinger i l'RSI estan més relacionats amb situacions de preu extrem o possible sobreventa, que poden deixar una forma més visible en el gràfic. En canvi, el MACD depèn del encreuament de mitjanes mòbils exponencials i de la seva línia de senyal. Aquest encreuament no sempre es veu de manera directa en una imatge de veles japoneses, perquè és el resultat d'un càlcul suavitzat sobre la sèrie de preus.

El sistema Hard Voting ofereix resultats estables, però no sempre supera el millor model individual. En les Bandes de Bollinger i en l'RSI queda per sota de la CNN. En el MACD obté una accuracy lleugerament superior, però el seu F1-score continua sent inferior al de la CNN. Això mostra que combinar models pot ajudar a reduir alguns errors, però no garanteix necessàriament superar el millor classificador individual.

4.3 Anàlisi de les matrius de confusió

Per entendre millor els resultats anteriors, també s'han revisat les matrius de confusió dels models principals. Aquesta anàlisi permet veure quants casos s'han classificat correctament i quins tipus d'errors apareixen en cada indicador.

La Taula 4 mostra els valors de veritats negatius (TN), falsos positius (FP), falsos negatius (FN) i veritats positius (TP) per als models més rellevants. En aquest treball, els falsos positius tenen una importància especial, ja que corresponen a casos en què el model prediu un senyal de compra, però la regla tècnica no el marca com a tal.

En les Bandes de Bollinger, la CNN presenta una matriu molt equilibrada. Només genera 11 falsos positius i 10 falsos negatius sobre 1.100 mostres de prova. Això indica que el model és capaç de reconèixer força bé tant les finestres

Taula 4: Matrius de confusió dels models principals en el conjunt de prova.

Indicador	Model	TN	FP	FN	TP
Bollinger	CNN	539	11	10	540
Bollinger	Random Forest	521	29	11	539
Bollinger	Hard Voting	524	26	10	540
MACD	CNN	392	158	73	477
MACD	Random Forest	421	129	102	448
MACD	Hard Voting	435	115	109	441
RSI	CNN	532	18	9	541
RSI	RBF SVM	530	20	13	537
RSI	Hard Voting	521	29	6	544

amb senyal com les finestres sense senyal. Random Forest i Hard Voting també obtenen resultats alts, però amb un nombre més gran de falsos positius.

En el cas de l'RSI també s'observa un comportament molt positiu. La CNN només comet 18 falsos positius i 9 falsos negatius. El model RBF SVM queda molt a prop, fet que mostra que aquest indicador es pot separar bé fins i tot amb models clàssics. Hard Voting redueix els falsos negatius fins a 6, però a canvi augmenta els falsos positius fins a 29.

El MACD és el cas més problemàtic. Tots els models generen més errors que en els altres dos indicadors. La CNN comet 158 falsos positius i 73 falsos negatius, mentre que Hard Voting redueix els falsos positius a 115, però augmenta els falsos negatius fins a 109. Això mostra que, en MACD, els models tenen més dificultats per decidir correctament quan hi ha un senyal de compra.

Aquesta diferència reforça la interpretació anterior. Les regles de l'RSI i de les Bandes de Bollinger semblen deixar patrons visuals més fàcils d'identificar en les imatges. En canvi, el MACD depèn de mitjanes mòbils exponencials i d'un encreuament entre línies suavitzades. Per això, el senyal pot no quedar representat de manera tan directa en la forma de les veles japoneses.

4.4 Anàlisi dels falsos positius

A més de comparar les mètriques generals, s'ha fet una revisió específica dels falsos positius. Aquest tipus d'error és especialment important en aquest treball, perquè representa els casos en què el model prediu un senyal de compra, però la regla tècnica utilitzada per generar l'etiqueta no el considera com a tal.

Dit d'una altra manera, un fals positiu és una finestra que el model interpreta com una possible oportunitat de compra, però que segons l'indicador tècnic correspon a la classe 0. En un escenari real, aquest error podria portar a considerar una entrada al mercat que no estaria justificada per la regla definida.

La Taula 5 mostra el nombre total de falsos positius obtingut pels diferents tipus de models. En els models clàssics s'han agregat totes les prediccions realitzades pels classificadors avaluats. En canvi, la CNN i el Hard Voting es mostren de manera separada, ja que corresponen als models finals més rellevants.

Els resultats mostren que el percentatge de falsos positius és relativament baix en tots els casos. Tot i això, el Hard Voting és el mètode que obté el percentatge més reduït, amb un 5,15%. La CNN també presenta un valor baix,

Taula 5: Falsos positius agregats segons el tipus de model.

Tipus de model	FP	Prediccions	Percentatge
Models clàssics	3.468	52.800	6,57%
CNN	187	3.300	5,67%
Hard Voting	170	3.300	5,15%

amb un 5,67%. Els models clàssics, considerats de manera agregada, arriben al 6,57%.

Aquesta comparació s'ha d'interpretar amb prudència, perquè el nombre de prediccions no és el mateix en tots els grups. Els models clàssics agrupen moltes més prediccions, ja que inclouen diversos classificadors. Tot i això, la taula permet veure que tant la CNN com el Hard Voting mantenen un nombre de falsos positius controlat.

Quan s'analitzen els falsos positius per indicador, s'observa que el MACD concentra la major part dels errors. En la CNN, dels 187 falsos positius totals, 158 provenen del MACD. En el Hard Voting passa una cosa similar: dels 170 falsos positius totals, 115 corresponen al MACD.

Taula 6: Falsos positius de la CNN i del Hard Voting per indicador.

Indicador	CNN	Hard Voting
Bandes de Bollinger	11	26
MACD	158	115
RSI	18	29

Aquesta diferència confirma que el MACD és l'indicador més difícil de classificar a partir de les imatges. En les Bandes de Bollinger i en l'RSI, els falsos positius són molt més baixos. Això pot indicar que aquests dos indicadors generen situacions visuals més recognoscibles, com moviments propers a zones de possible sobreventa.

En canvi, el MACD depèn del encreuament entre dues línies calculades a partir de mitjanes mòbils exponencials. Aquest encreuament pot no aparèixer de manera clara en la forma de les veles japoneses. Per aquest motiu, el model pot detectar patrons que visualment semblen una possible compra, però que no compleixen exactament la regla algebraica del MACD.

Per complementar aquesta anàlisi, a l'Apèndix B es mostren exemples visuals de veritables positius i falsos positius obtinguts per la CNN en el conjunt de prova. Aquests exemples no s'inclouen per fer una anàlisi individual de cada ticker, sinó per ajudar a interpretar millor què pot estar aprenent el model a partir de la forma de les veles japoneses.

En alguns casos, els falsos positius no semblen errors completament arbitraris. És a dir, el model pot predir una finestra com a positiva perquè visualment presenta característiques semblants a les d'un senyal real, com una caiguda recent del preu, un possible canvi de direcció o una zona de debilitat prèvia. Tot i això, aquestes finestres no compleixen exactament la regla algebraica utilitzada per generar l'etiqueta. Això és especialment important en el MACD, ja que el senyal depèn d'un encreuament entre mitjanes mòbils exponencials i no sempre es pot observar directament només mirant la forma de les veles.

Per tant, l'anàlisi dels falsos positius reforça una de les conclusions principals del treball: els models poden recuperar força bé algunes regles tècniques a partir d'imatges, però no totes les regles tenen la mateixa traducció visual.

Les Bandes de Bollinger i l'RSI semblen més fàcils d'identificar visualment, mentre que el MACD presenta més ambigüitat.

4.5 Avaluació borsària simplificada

A més de les mètriques de classificació, s'ha fet una avaluació borsària simplificada a partir de les prediccions de la CNN. L'objectiu d'aquesta prova no és construir una estratègia de trading completa, sinó observar si els senyals generats pel model tindrien un comportament raonable quan s'apliquen sobre dades no balancejades.

Aquesta distinció és important. Els models s'han entrenat i avaluat amb conjunts balancejats per poder comparar correctament les classes 0 i 1. En canvi, en un context real els senyals de compra són molt menys freqüents que els dies sense senyal. Per això, aquesta prova complementària s'ha aplicat sobre dades no balancejades, més properes a la distribució real del mercat.

La regla utilitzada és senzilla. Quan el model prediu un senyal de compra, es considera una entrada a l'obertura de l'endemà. La sortida es fa al tancament cinc sessions després. Amb aquesta regla s'han calculat diferents mètriques borsàries, com el retorn total, el CAGR, el Sharpe Ratio, el Sortino Ratio i el Maximum Drawdown.

Taula 7: Resultats de l'avaluació borsària simplificada amb les prediccions de la CNN.

Mètrica	Tickers del projecte	Tickers externs
CAGR	0,3544	0,2658
Sharpe Ratio	2,5355	1,8311
Sortino Ratio	4,0614	2,7123
Maximum Drawdown	-0,2398	-0,2988
Retorn total	3,5580	2,2493
Operacions	23,053	9,882

Els resultats mostren un comportament positiu tant en els tickers utilitzats dins del projecte com en els tickers externs. En els tickers del projecte s'obté un CAGR de 0,3544 i un Sharpe Ratio de 2,5355, mentre que en els tickers externs aquests valors són de 0,2658 i 1,8311, respectivament.

Aquesta diferència indica que el comportament és millor en els actius utilitzats durant el desenvolupament del projecte, però també mostra que el model manté resultats positius quan s'aplica sobre tickers no utilitzats en la construcció principal del treball. Tot i això, aquesta prova s'ha d'interpretar amb prudència, ja que es tracta d'una avaluació simplificada. No incorpora elements com comissions, lliscament, liquidat o una gestió completa del capital.

Tot i aquestes limitacions, l'avaluació borsària aporta informació útil. Les mètriques de classificació indiquen si el model és capaç d'entendre les regles tècniques. En canvi, aquesta prova permet observar què passaria si les prediccions es transformessin en operacions senzilles. Així, es complementa l'anàlisi del model amb una visió més propera a l'ús pràctic dels senyals.

4.6 Síntesi dels resultats

En conjunt, els resultats mostren que la representació visual de les sèries financeres permet recuperar senyals tècnics de compra amb un bon rendiment, especialment en les Bandes de Bollinger i en l'RSI. En aquests dos indicadors, la CNN

obté resultats molt alts i els models clàssics també són competius.

El MACD és l'indicador més difícil. Tant les mètriques generals com les matrius de confusió i l'anàlisi dels falsos positius mostren que aquest indicador genera més errors. La causa principal és que la seva regla depèn de mitjanes mòbils exponencials i d'un encreuament entre línies suavitzades, un patró que no sempre queda reflectit de manera directa en la forma de les veles japoneses.

També s'observa que el Hard Voting dona resultats estables, però no sempre millora el millor model individual. Això indica que combinar classificadors pot ajudar en alguns casos, però no garanteix superar una CNN ben ajustada. Finalment, l'avaluació borsària simplificada mostra que les prediccions del model poden tenir un comportament positiu en dades no balancejades, tot i que aquesta prova s'ha d'entendre com una anàlisi complementària i no com una estratègia de trading completa.

5 CONCLUSIONS I TREBALL FUTUR

Aquest treball ha tingut com a objectiu estudiar si és possible classificar senyals tècnics de compra a partir d'imatges de veles japoneses generades amb dades financeres històriques. Per fer-ho, s'ha construït un procés complet que parteix de dades OHLC de l'S&P 500, calcula indicadors tècnics, genera etiquetes binàries, transforma finestres temporals en imatges i compara diferents models de classificació.

Els resultats obtinguts mostren que la representació visual de les sèries financeres pot ser útil per recuperar algunes regles tècniques de compra. En concret, les Bandes de Bollinger i l'RSI han obtingut resultats molt elevats, especialment amb la CNN. Això indica que aquests indicadors generen patrons que poden quedar força ben reflectits en la forma de les veles japoneses.

El MACD, en canvi, ha estat l'indicador més difícil de classificar. Tant les mètriques generals com les matrius de confusió i l'anàlisi dels falsos positius mostren que aquest indicador produeix més errors. Aquesta diferència sembla relacionada amb la definició del MACD, ja que depèn de mitjanes mòbils exponencials i d'un encreuament entre línies suavitzades. Aquest tipus de senyal no sempre es veu de manera directa en una imatge de veles.

També s'ha observat que els models clàssics continuen sent competius en alguns casos. Tot i que la CNN ha obtingut els millors resultats globals en termes d'F1-score, models com Random Forest o RBF SVM han aconseguit rendiments molt propers en alguns indicadors. Per tant, el treball no mostra només que la CNN funciona bé amb imatges, sinó també que una representació visual senzilla pot ser aprofitada per diferents tipus de classificadors.

L'anàlisi dels falsos positius ha permès entendre millor els errors dels models. Aquesta part és important perquè, en un context financer, no n'hi ha prou amb tenir bones mètriques agregades. També cal saber quan el model pot generar un senyal de compra que no queda justificat per la regla tècnica utilitzada. En aquest sentit, el MACD ha concentrat la major part dels falsos positius, reforçant la idea que és l'indicador amb una traducció visual més complexa.

Com a complement, s'ha fet una avaluació borsària simplificada amb les prediccions de la CNN sobre dades no ba-

lancejades. Els resultats han estat positius tant en els tickers del projecte com en tickers externs. Tot i això, aquesta prova s'ha d'interpretar amb prudència, ja que no inclou comissions, lliscament de preus ni una gestió completa del capital. Per tant, no es pot considerar una estratègia de trading definitiva, sinó una anàlisi addicional per observar el comportament pràctic dels senyals.

En conjunt, el treball mostra que transformar sèries temporals financeres en imatges pot ser una aproximació útil per classificar senyals tècnics. Així i tot, els resultats també indiquen que no totes les regles tècniques són igualment fàcils de recuperar visualment. Com a línies futures, es podria ampliar l'estudi amb més indicadors, comparar diferents mides de finestra, provar arquitectures de xarxes neuronals més avançades i desenvolupar un backtesting més complet que incorpori costos de transacció i gestió de capital.

ÚS DE LA IA GENERATIVA

Durant la realització d'aquest projecte s'ha utilitzat intel·ligència artificial generativa, concretament ChatGPT, com a eina de suport per a la revisió de la redacció, la correcció lingüística i la millora de la claredat expositiva del document. Aquest ús s'ha limitat a aspectes formals del text i no a la generació del contingut tècnic, metodològic o analític del treball.

AGRAÏMENTS

Vull agrair al meu tutor, Albert Romero Sanchez, el seu seguiment i les orientacions rebudes durant el desenvolupament d'aquest treball. Els seus comentaris han estat útils per revisar l'enfocament del projecte i millorar-ne la metodologia i la presentació.

Finalment, agraeixo a l'Escola d'Enginyeria de l'Universitat Autònoma de Barcelona la formació rebuda durant aquests anys, que ha fet possible desenvolupar aquest treball.

REFERÈNCIES

- [1] N. Cohen, T. Balch, and M. Veloso, "Trading via Image Classification," in *Proceedings of the First ACM International Conference on AI in Finance (ICAIF '20)*, New York, NY, USA, 2020. doi: <https://doi.org/10.1145/3383455.3422544>.
- [2] J.-H. Chen and Y.-C. Tsai, "Encoding Candles-ticks as Images for Patterns Classification Using Convolutional Neural Networks," *arXiv preprint arXiv:1901.05237*, 2019. Available: <https://arxiv.org/abs/1901.05237>.
- [3] R. Aroussi, "yfinance: Download market data from Yahoo! Finance's API," GitHub repository. Available: <https://github.com/ranaroussi/yfinance>.
- [4] R. Aroussi, "yfinance documentation." Available: <https://ranaroussi.github.io/yfinance/>.

- [5] W. McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference*, Austin, TX, USA, 2010, pp. 56–61. doi: <https://doi.org/10.25080/Majora-92bf1922-00a>.
- [6] C. R. Harris, K. J. Millman, S. J. van der Walt, et al., "Array programming with NumPy," *Nature*, vol. 585, pp. 357–362, 2020. doi: <https://doi.org/10.1038/s41586-020-2649-2>.
- [7] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. Available: <https://jmlr.org/papers/v12/pedregosa11a.html>.
- [8] Scikit-learn developers, "Scikit-learn User Guide." Available: https://scikit-learn.org/stable/user_guide.html.
- [9] A. Paszke, S. Gross, F. Massa, et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems*, vol. 32, 2019. Available: <https://arxiv.org/abs/1912.01703>.
- [10] PyTorch, "PyTorch documentation." Available: <https://docs.pytorch.org/>.
- [11] Matplotlib, "mplfinance: Financial Markets Data Visualization using Matplotlib," GitHub repository. Available: <https://github.com/matplotlib/mplfinance>.
- [12] Pillow contributors, "Pillow documentation." Available: <https://pillow.readthedocs.io/>.
- [13] Investopedia, "Relative Strength Index (RSI): What It Is, How It Works, and Formula." Available: <https://www.investopedia.com/terms/r/rsi.asp>.
- [14] Investopedia, "Moving Average Convergence Divergence (MACD)." Available: <https://www.investopedia.com/terms/m/macd.asp>.
- [15] Investopedia, "Bollinger Bands." Available: <https://www.investopedia.com/terms/b/bollingerbands.asp>.
- [16] Scikit-learn developers, "Metrics and scoring: quantifying the quality of predictions." Available: https://scikit-learn.org/stable/modules/model_evaluation.html.
- [17] Scikit-learn developers, "f1_score." Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html.
- [18] Scikit-learn developers, "balanced_accuracy_score." Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.balanced_accuracy_score.html.
- [19] Scikit-learn developers, "matthews_corrcoef." Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.matthews_corrcoef.html.
- [20] Scikit-learn developers, "confusion_matrix." Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html.
- [21] Investopedia, "Sharpe Ratio: Definition, Formula, and Examples." Available: <https://www.investopedia.com/terms/s/sharperatio.asp>.
- [22] Investopedia, "Sortino Ratio: Definition, Formula, Calculation, and Example." Available: <https://www.investopedia.com/terms/s/sortinoratio.asp>.
- [23] Investopedia, "Understanding Drawdowns: Definition, Risks, and Recovery Time." Available: <https://www.investopedia.com/terms/d/drawdown.asp>.

APÈNDIX A. RÀNQUING RESUMIT DELS MODELS AVALUATS

Aquest apèndix mostra un rànkung resumit dels models amb millor rendiment en el conjunt de prova per a cada indicador tècnic. Les files estan ordenades segons l'F1-score. L'objectiu és complementar la Taula 3 i mostrar que els models destacats en l'anàlisi principal provenen d'una comparació més àmplia entre diferents classificadors.

Taula A1: Models amb millor F1-score per indicador en el conjunt de prova.

Model	Accuracy	F1	MCC
Bandes de Bollinger			
CNN	0,9809	0,9809	0,9618
Hard Voting	0,9673	0,9677	0,9349
Random Forest	0,9636	0,9642	0,9278
Extra Trees	0,9627	0,9633	0,9259
RBF SVM	0,9618	0,9620	0,9237
MACD			
CNN	0,7900	0,8051	0,5871
Random Forest	0,7955	0,7993	0,5913
Bagging	0,7927	0,7986	0,5864
Extra Trees	0,7945	0,7986	0,5896
Hard Voting	0,7964	0,7975	0,5928
RSI			
CNN	0,9755	0,9757	0,9510
RBF SVM	0,9700	0,9702	0,9401
Hard Voting	0,9682	0,9688	0,9372
Deep Neural Net	0,9673	0,9675	0,9346
Extra Trees	0,9636	0,9645	0,9285

APÈNDIX B. EXEMPLES VISUALS DE CLASIFICACIÓ

Aquest apèndix recull els exemples visuals utilitzats per complementar l'anàlisi dels falsos positius presentada a l'apartat 4.4. Les mostres provenen del conjunt de prova i corresponen a prediccions de la CNN. Per a cada indicador es mostra un verdader positiu i un fals positiu.

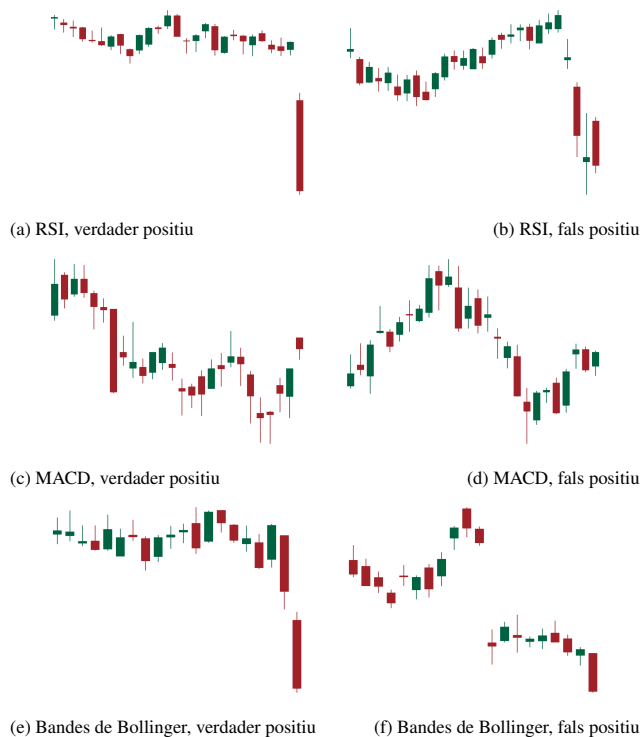


Fig. B1: Exemples visuals de verdaers positius i falsos positius obtinguts per la CNN en el conjunt de prova.

APÈNDIX C. MOSTRES UTILITZADES EN ELS EXEMPLES VISUALS

Aquest apèndix mostra la informació bàsica de les finestres utilitzades en els exemples visuals de l'Apèndix B. Totes les mostres provenen del conjunt de prova i corresponen a prediccions realitzades per la CNN. En la taula, VP indica verdader positiu i FP indica fals positiu.

Taula C1: Finestres utilitzades en els exemples visuals de classificació.

Indic.	Cas	Ticker	Període	Real/Pred.	Dies
RSI	VP	SJM	02/05/25–10/06/25	1/1	27
RSI	FP	GL	03/03/25–08/04/25	0/1	27
MACD	VP	IBM	10/12/24–17/01/25	1/1	26
MACD	FP	PRU	18/06/25–25/07/25	0/1	26
Boll.	VP	AVB	03/07/25–31/07/25	1/1	20
Boll.	FP	DELL	12/11/24–10/12/24	0/1	20