

UNIVERSITÀ DEGLI STUDI DI MILANO
Facoltà di Scienze e Tecnologie
Corso di Laurea in informatica

PROGETTAZIONE E SVILUPPO DI
TEST SUITE AUTOMATICA E
MODULO RACCOLTA DATI PER
INTELLIGENZE ARTIFICIALI CHE SI
OCCUPANO DI CREARE STRATEGIE
DI INVESTIMENTO

Relatore: Prof. Alfio Ferrara

Correlatore: Dr. Andrea Bellacicca

Tesi di:
Garion Musetta
Matricola: 920623

Anno Accademico 2019-2020

dedicato a ...

Prefazione

Organizzazione della tesi

La tesi è organizzata come segue:

- nel Capitolo 1

Ringraziamenti

Indice

	ii
Prefazione	iii
Ringraziamenti	iv
0.1 Introduzione	1
1 Financial trading	2
1.1 Criptovalute	3
1.2 Modello dei dati	4
1.3 Guadagnare con il trading	8
1.4 Esempi di strategie	11
1.4.1 Buy-Hold	12
1.4.2 Moving Average Crossover	12
1.5 Ruolo delle intelligenze artificiali e scopo della tesi	14
2 AI per financial prediction e Sentymment	16
2.1 Stato dell'arte delle AI	17
2.1.1 Supporto di informazioni relative agli eventi	19
2.1.2 Punti chiave nelle predizioni finanziarie	20
2.1.3 Metodi di previsione	22
2.2 Sentymment	34
2.2.1 Cenni all'architettura	34
2.2.2 Strumenti e tecnologie utilizzati	37
3 Elaborazione dati di trading	40
3.1 Download dati Kraken	40
3.2 Creazione candele OHLCV	45
3.3 Indicatori	46
3.4 Creazione strategie	46

4	Test per intelligenze artificiali	47
4.1	Test per decision support AI	49
4.2	Test framework workflow	49
4.3	Analisi offline: maximum profit e genetic AI	49
4.4	Analisi online: reinforcement learning	49
4.4.1	Q-learning	51
4.4.2	Multi-armed bandit	51
5	Conclusioni e lavori futuri	52

0.1 Introduzione

Scopo di questo documento è affrontare il tema delle intelligenze artificiali per trading, illustrare alcune tecniche, applicare delle metodologie di test per analizzarne le prestazioni e descrivere lo sviluppo di nuovi componenti da integrare nelle già esistenti intelligenze artificiali (AI).

Sono disponibili attualmente una grande varietà di algoritmi di AI nel campo della finanza, in particolare nella previsione del mercato azionario, in grado di produrre ottimi risultati. Qualsiasi sia la tecnica utilizzata, quando si sviluppa uno strumento di questo tipo è indispensabile porsi delle domande circa la qualità dei risultati prodotti, confrontandoli con un modello ottimale di riferimento. Spesso però questo modello non è disponibile oppure è molto difficile da creare, a volte anche incalcolabile. I risultati dovrebbero anche essere paragonati con quelli prodotti da altre applicazioni esistenti che operano nel campo in analisi.

Viene considerato uno fra i numerosi sistemi di AI per trading, *Sentyment*. Il software è sottoposto ad analisi e test e vengono aggiunti nuovi componenti: un modulo di raccolta dati e un'ulteriore livello di AI in grado di apprendere l'andamento dell'algoritmo e di conseguenza aggiornare alcuni suoi parametri, in automatico.

Capitolo 1

Financial trading

Il dominio considerato è quello della finanza, in particolare lo sviluppo di strategie di investimento intelligenti.

L'intelligenza artificiale da testare, *Sentyment*, è un software commerciale che genera diverse strategie di investimento personalizzate per numerosi strumenti finanziari ed è anche in grado di piazzare direttamente ordini sul mercato, cioè effettuare operazioni di acquisto e vendita. Il mercato in cui opera è quello delle criptovalute. Di queste, Bitcoin è il tipo più famoso ed asset più valevole a livello di borsa valori, insieme ad altre monete come Ethereum, Bitcoin Cash e Ripple, meno famose ma con le loro particolarità e altrettanto scambiate.

L'obiettivo di Sentyment dunque non è esattamente la previsione dell'andamento di un certo titolo azionario. Non è necessario predire con precisione assoluta il valore di un titolo in un dato istante di tempo per massimizzare i guadagni. Ciò su cui si focalizza è diversificare il portafoglio e adattarsi alle esigenze degli utenti. Il suo scopo è produrre delle *strategie di investimento*, ossia approcci di investimento personalizzati in funzione della propensione al rischio, degli obiettivi e degli interessi specifici del singolo investitore. Seguendo tale strategia, l'investitore può decidere tra diversi tipi di attività da includere nel proprio portafoglio di investimento.

Una specifica strategia di investimento può essere determinata da una serie di fattori, tra cui la propensione al rischio e i rendimenti che si vogliono perseguire sugli investimenti, nonché le attività, le regioni e i settori a cui si è interessati e il periodo per il quale si intende investire.

Definita una strategia, Sentyment la persegue selezionando quindi, in ogni "istante di tempo" un'operazione tra: *buy* (comprare il titolo azionario), *sell* (vendere) o *hold* (mantenere il portafoglio). Lo scopo a lungo termine della strategia è massimizzare il guadagno rispetto a quanto investito, oltre a rispettare i vincoli di rischio e interessi

specifici dell'investitore. La AI è in grado di decidere quali fra queste operazioni effettuare al fine di massimizzare gli obiettivi.

L'istante di tempo usato da Sentymment è l'ora. Ogni ora la piattaforma di trading di riferimento produce nuovi dati e li rende disponibili per essere scaricati dai trader, che possono quindi compiere le azioni sopra descritte. Nel caso si volesse cambiare il periodo di calcolo è anche possibile scegliere i minuti: 1, 5, 15, 30, 60, 240, 1440, 10080, 21600.

In questo capitolo sono spiegati alcuni dei concetti e delle terminologie del campo della finanza utili a comprendere il lavoro svolto dalle AI per trading.

1.1 Criptovalute

Sentymment si concentra sul mercato delle cryptocurrency. Ci sono alcuni vantaggi che questo mercato offre rispetto allo scambio di valute tradizionali come Euro-Dollaro.

Criptovaluta si riferisce ad una rappresentazione digitale di valore basata sulla crittografia. Utilizzano tecnologie di tipo peer-to-peer su reti i cui nodi risultano costituiti da computer di utenti, situati potenzialmente in tutto il globo. Su questi computer vengono eseguiti appositi programmi che svolgono funzioni di portamonete. Non c'è attualmente alcuna autorità centrale che le controlla. Le transazioni e il rilascio avvengono collettivamente in rete, previo raggiungimento di un consenso distribuito fra gli utenti.

Le criptovalute facilitano lo scambio di valore e informazioni tra due parti. Lo fanno in modo sicuro e affidabile, in modo che non sia necessaria una terza parte per controllare e approvare la transazione. Dal punto di vista dell'utente, le operazioni di scambio avvengono tra il suo software e la *piattaforma di exchange*. Una piattaforma di trading (o exchange) è un software utilizzato per il trading: permette apertura, chiusura e gestione delle posizioni di mercato attraverso un intermediario finanziario come un broker online. Nel caso di Sentymment, la piattaforma di riferimento è Kraken¹.

Le criptovalute esistono soltanto all'interno delle strutture Blockchain, o qualsiasi sia la forma di consenso sottostante la moneta, quindi si rivelano particolarmente adatte all'utilizzo da parte di *bot* o programmi. Non si deve possedere fisicamente la moneta per effettuare scambi, basta un wallet associato alla piattaforma di exchange e il bot è libero di operare sul mercato.

La differenza principale rispetto alle valute tradizionali è l'assenza di un'autorità

¹Kraken Digital Asset Exchange. URL: <https://www.kraken.com/>

centrale che controlli il rilascio della moneta. Questo facilita il lavoro dei trader di cryptocurrency, che devono fare riferimento soltanto alla piattaforma di exchange.

La criptovaluta conosciuta da tutti è Bitcoin. Ne esistono però diverse decine e, se si considerano le coppie di scambio (esempio, Bitcoin-Ethereum), sono quasi cento quelle trattate da Kraken. Al fine della diversificazione del portafoglio, che consiste in una riduzione della rischiosità del rendimento, legata alla presenza di più attività finanziarie, i cui rendimenti non sono correlati, è opportuno investire in più asset possibili per avere risultati migliori in termini di guadagno.

Alcune fra le monete trattate da Kraken (e Sentyment) sono:

- BTC - Bitcoin, la prima più famosa e col valore più alto.
- ETH - Ethereum è una piattaforma decentralizzata del per la creazione e pubblicazione peer-to-peer di contratti intelligenti creati in un linguaggio di programmazione Turing-completo.
- ZEC - Zcash offre privacy e trasparenza selettiva delle transazioni. I pagamenti sono pubblicati su una blockchain pubblica, ma il mittente, il ricevente e il valore della transazione possono rimanere privati.
- BCH - Bitcoin Cash è un fork di Bitcoin.
- XRP - Ripple

1.2 Modello dei dati

I dati usati dalla AI in questione sono scaricati dalla piattaforma di trading **Kraken**, usata per lo scambio di criptovalute. Ogni secondo, trader da tutto il mondo effettuano azioni di compravendita di titoli, facendo crescere o diminuire il valore azionario di ogni asset, esattamente nel modo in cui opera anche Sentyment. Kraken rende disponibile, attraverso delle API (application programming interface), un elenco di record contenenti timestamp e prezzo, che formano lo storico degli scambi effettuati per ognuno degli asset che espone. Questi sono 110 coppie di valute che rappresentano i valori di scambio fra criptovalute e USD / EUR. Si fa distinzione fra dati "RAW", ovvero un semplice elenco di prezzi variabili nel tempo, e le "candele OHLCV" (*open, high, low, close*).

Le candele sono uno degli strumenti grafici più popolari in quanto offrono un eccellente riferimento virtuale dei movimenti dei prezzi in un intervallo di tempo: al

minuto, all'ora, al giorno, mensile ed altro.

Sono dunque la rappresentazione di un dato aggregato secondo l'unità di tempo selezionata. Le informazioni che racchiudono sono indicate dal loro nome. Prendendo come esempio le candele orarie: la creazione di ognuna di esse parte allo scoccare dell'ora e fissa un prezzo di apertura (*open*: il prezzo di vendita dell'asset in quel momento), per terminare dopo un'ora con un prezzo di chiusura (*close*: lo stesso prezzo di vendita dell'asset al momento di chiusura, che sarà ora cambiato rispetto a *open*); si calcola quali sono stati i picchi massimi e minimi di prezzo durante l'ora (*high*, *low*) e infine il volume, che rappresenta l'ammontare totale scambiato nell'ora. Le candele OHLCV rappresentano dunque la storia dell'andamento dei prezzi di un certo asset durante il periodo di tempo fissato: iniziano con un certo prezzo, all'apertura dello slot temporale, che durante il periodo considerato evolve e crea i punti minimo e massimo, per poi terminare alla chiusura dello slot con la determinazione del prezzo finale; i prezzi minimo e massimo possono superare anche di molto quelli di apertura e chiusura. Esiste la possibilità per cui il prezzo non evolva nel periodo selezionato e che i quattro termini, *open*, *close*, *high* e *low*, coincidano. Volume è invece la somma totale della quantità di titoli comprati e venduti durante l'unità di tempo.

Sentymment lavora quasi esclusivamente su queste candele e, anche se le API di Kraken permettono di scaricare dati già aggregati, la AI scarica soltanto dati di trading raw per poi creare autonomamente le sue candele attraverso un modulo dati dedicato.

Ultimo Aggiornamento: 02-05-20 15:06:40 +00:00

Trade recenti (XBT/EUR)

1 - 15 di 250 trade

Ora ▼	Ordine ↕	Prezzo ↕	Volume ↕
15:06:36 +00:00	sell	€8,567.9	0.01946901
15:06:21 +00:00	buy	€8,568.0	0.02432718
15:06:17 +00:00	sell	€8,567.9	0.00453099
15:06:17 +00:00	sell	€8,567.9	0.01546901
15:06:04 +00:00	sell	€8,567.9	0.00288574
15:06:03 +00:00	sell	€8,567.9	0.03500000
15:06:01 +00:00	buy	€8,568.0	0.01082337
15:05:58 +00:00	buy	€8,568.0	0.05478300
15:05:57 +00:00	buy	€8,568.0	0.00673553
15:05:55 +00:00	buy	€8,568.0	0.04054651
15:05:54 +00:00	buy	€8,568.0	0.04081193
15:05:48 +00:00	buy	€8,568.0	0.06738600
15:05:45 +00:00	buy	€8,568.0	0.16574437

Figura: Elenco transazioni Bitcoin/Euro relative ad una finestra di minuti. I record mostrano l'ora in cui è avvenuto il trade, il tipo di operazione (buy / sell), il prezzo di scambio del titolo e la quantità di titoli scambiati. L'elenco dei dati di trade 'raw' è disponibile tramite le API kraken ed è la fonte grezza di dati finanziari utilizzati per calcolare i grafici OHLCV, fondamentali per l'analisi dei mercati. (fonte: <https://www.kraken.com/>)



Figura: Grafico OHLCV ricavato dai prezzi in Figura 1. Sulle ascisse è rappresentata l'ora mentre le ordinate sono il prezzo (in Euro) del titolo Bitcoin. Le "barre" orizzontali verdi e rosse sono le candele OHLCV: verdi se il prezzo è in crescita (*open* minore di *close*), rosse se in discesa (*open* maggiore di *close*). Le candele sono di durata 1 ora e quindi i prezzi elencati nella precedente immagine rientrano soltanto in parte nell'ultima candela (rossa) delle 15:00 - 16:00, ancora aperta e quindi in creazione. La candela è rossa perchè, come si nota dai prezzi, il valore del titolo è in discesa: partendo da circa 8.568 (probabilmente più alto nei record precedenti) si scende verso 8.567.

Le linee colorate rappresentano le medie mobili dei prezzi di chiusura; usate per analisi tecnica, sono rispettivamente le medie a 10 (verde), 21 (azzurro) e 100 (blu) candele. Più candele sono considerate nella media e meno questa cambierà bruscamente, avendo quindi la media a 100 candele (cioè 100 ore) molto più lenta delle altre. (fonte:

<https://www.kraken.com/>)

Le immagini descrivono l'andamento dei prezzi scambio per scambio e le corrispondenti candele ohlcv come esposte dalla piattaforma di trading. I bordi orizzontali che compongono i lati superiore e inferiore della candela sono il prezzo di apertura e quello di chiusura: se la candela è rossa il prezzo di chiusura è minore di quello di apertura (significa che il prezzo dell'asset è sceso durante l'unità di tempo), e quindi il bordo superiore rappresenta il valore di *open* e quello inferiore *close*, mentre una candela verde indica una crescita del prezzo (*open* è il lato inferiore mentre *close* quello superiore). Le linee verticali della candela si estendono invece verso il prezzo minimo e massimo che l'asset ha toccato nel periodo di tempo fra apertura e chiusura della candela.

Le API di Kraken permettono di effettuare diversi tipi di operazioni, come scaricare lo storico dei prezzi di tutte le coppie valuta-criptoaluta disponibili, dal 2013; restare in attesa per ricevere in tempo reale i nuovi dati sugli scambi di titoli; inoltre è anche possibile creare il proprio portafogli sulla piattaforma e piazzare ordini di acquisto e vendita. Tutte funzionalità ben sfruttate da Sentyment, che opera sul mercato in autonomia.

1.3 Gudagnare con il trading

E' importante notare che la ricchezza in possesso, calcolata in ogni istante, è data dalla quantità di budget e dalla quantità di titoli posseduti. Gli asset finanziari hanno un valore monetario indicato dal loro prezzo, quindi il possesso di alcuni di questi, anche se il budget da investire è azzerato, comporta comunque una ricchezza investita che si può recuperare rivendendo il titolo.

Si prende come esempio l'asset XBT/EUR, valuta di scambio fra Bitcoin e Euro. Partendo da un budget iniziale in Euro, bisogna decidere quando comprare titoli di Bitcoin e quando rivenderli per avere di nuovo Euro. Lo scopo della AI, quindi, non è soltanto massimizzare gli Euro in possesso, ma anche il numero di titoli Bitcoin o del quantitativo di entrambi i titoli da avere in portafoglio che porti all'aumento del loro valore.

A seconda delle aspettative sul rialzo o ribasso di un titolo è possibile agire in due modi: andare *long* o andare *short*. Andare **long** vuol dire acquistare Bitcoin vendendo Euro, puntando ad un apprezzamento (aumento di valore) della criptoaluta; in questo caso l'aspettativa è di un aumento del valore del cambio e si avrà quindi una visione rialzista (bullish) del mercato. Andare **short** vuol dire vendere Bitcoin acquistando Euro, puntando quindi un apprezzamento dell'Euro nei confronti della criptomoneta. In questo caso l'aspettativa è di una diminuzione di valore del cambio.

Si aprirà quindi una posizione che mira a una flessione del prezzo. La visione che il trader avrà del mercato sarà ribassista (bearish).

Il punto centrale su cui si basano le strategie di investimento è indovinare se la valuta aumenterà o diminuirà di prezzo, per poi andare long o short a seconda della previsione. L'**analisi tecnica** è un metodo riconosciuto e largamente utilizzato per anticipare o prevedere l'andamento dei prezzi e si basa sull'aspetto tecnico del mercato, utilizzando grafici e dati storici. In particolare vengono analizzati prezzi, volumi scambiati e fasce temporali. Si utilizzano indicatori matematici, statistici e grafici e i segnali e i suggerimenti ricavati da questi strumenti accompagnano i trader nelle loro decisioni in merito all'apertura e chiusura di posizioni di trading. L'analisi tecnica è un metodo di previsione dei prezzi basato sui dati storici.

Supponendo che il mercato ad un certo punto cresca e si abbia "indovinato" un buon numero di previsioni, ci si ritroverà a possedere dei titoli che ora valgono un prezzo superiore rispetto al loro valore iniziale. Se i titoli crescono di valore, è bene quindi acquistarne finché salgono, in questo modo si possiede un maggior numero di asset che andranno a valere sempre di più; in caso di perdita di valore del titolo, invece, generalmente si dovrebbe vendere i titoli, per riacquistare del budget investito che ora stava perdendo valore e poterlo investire in altri titoli che crescono.

Quando si decide di acquistare si sta effettuando la seguente operazione: dato un certo budget iniziale *budget*, il valore del titolo *price* e la quantità di titolo acquistato *equity* e supponendo di spendere sempre tutto il budget per acquistare il titolo

$$equity = budget / price \quad (1)$$

Mentre, all'occorrenza di un'operazione di vendita, utilizzando *equity* appena acquistata:

$$budget = equity * price \quad (2)$$

È possibile in ogni momento calcolare il ricavo o quantità di beni in possesso combinando il budget con *equity*, tenendo presente che dopo un acquisto il budget scende a zero e *equity* assume il valore indicato, mentre dopo una vendita è *equity* a scendere a zero e si ritorna in possesso di budget. Lo stesso budget riacquistato verrà usato nuovamente per comprare dei titoli facendo così crescere di nuovo *equity*, e così via.

Un vincolo è quello di non poter effettuare due medesime operazioni di fila.

Se si considera un certo budget iniziale *initial_budget*, il **guadagno** (*gain*) in ogni istante è dato da:

$$gain = (budget + equity * price) - initial_budget \quad (3)$$

Questa è la formula del guadagno che verrà usata da qui in avanti.

A questo punto è necessario introdurre il concetto di **fee**.

Le varie piattaforme di trading addebitano un costo per ogni singola operazione, sia di acquisto che di vendita, intestandosi una percentuale di budget speso dai trader come tassa per mantenere la piattaforma stessa. Nel caso di operazioni di acquisto, la tassa è calcolata come percentuale del budget speso per comprare il titolo, quindi viene scalata subito e il risultato è l'acquisto di una quantità leggermente minore di titoli (la restante dei quali viene pagata alla piattaforma); per le vendite invece la fee è tolta dal budget una volta che questo è ricalcolato vendendo equity, risultando in un ricavo leggermente minore di quello che si sarebbe ottenuto vendendo realmente l'intera quantità di titoli.

Nel caso di Kraken, la piattaforma di trading di riferimento su cui opera Sentyment, le fee sono del 0.26%: significa che viene sottratto il 0.26% di budget che si intende investire, per ogni operazione. Per altri mercati, valute o piattaforme di trading le fee possono variare e potrebbero non essere basate su una percentuale di acquisto ma fisse (flat fee vs per share fee).

La presenza delle fee modifica abbastanza drasticamente il funzionamento del mercato e l'efficacia delle strategie di investimento, che devono essere adattate ad un mercato con fee; è necessario quindi riscrivere le formule per il guadagno e per le operazioni di acquisto e vendita titoli, che ora comprendono la percentuale di fee.

Per gli acquisti, considerando un certo budget iniziale *budget* (o il budget risultato di una precedente vendita), il valore del titolo *price*, la quantità di titolo acquistato *equity* e una tassa *fee*:

$$equity = (budget * (1 - fee)) / price \quad (4)$$

Per le vendite:

$$budget = (equity * price) * (1 - fee) \quad (5)$$

Dal momento che le budget e equity sono già calcolati togliendo le fee per ogni operazione di acquisto e vendita, la formula del guadagno rimane la stessa.

Senza addentrarsi nelle diversità dei titoli e dei mercati, si può affermare che una strategia generale e semplificata per guadagnare facendo trading è *compra basso, vendi alto*; nella realtà i segnali di acquisto possono essere dati da altri indicatori più complessi e non sempre è possibile capire quando un titolo sta toccando un prezzo particolarmente alto o basso: è facile analizzare i grafici a posteriori, ma non altrettanto semplice prevedere massimi e minimi locali online. Per questo si utilizzano delle semplici strategie fondamentali che aiutano a capire l'andamento dei prezzi.

1.4 Esempi di strategie

Come già accennato, l'**analisi tecnica** è lo studio dell'andamento dei prezzi dei mercati finanziari nel tempo, allo scopo di prevederne le tendenze future, mediante principalmente metodi grafici e statistici. In senso lato è quella teoria di analisi secondo cui è possibile prevedere l'andamento futuro del prezzo di un bene quotato, studiando la sua storia passata. Viene utilizzata, assieme all'analisi fondamentale, per la definizione delle decisioni di operatività finanziaria.

L'analisi tecnica si prefigge di analizzare e comprendere, attraverso l'analisi del grafico, l'andamento dei prezzi, il quale a sua volta rispecchia le decisioni degli investitori; inoltre si basa sull'assunto fondamentale che, poiché il comportamento degli investitori si ripete nel tempo, al verificarsi di certe condizioni grafiche, anche i prezzi si muoveranno di conseguenza. Il compito principale dell'analisi tecnica è quindi quello dell'identificare un cambiamento di tendenza rispetto ad uno stadio iniziale, mantenendo una posizione di investimento fino a quando non vi sia prova che la tendenza stessa si sia di nuovo invertita.

Per individuare un trend si fa uso di indicatori tecnici, basati su prezzo e volume, calcolati a partire dai grafici: questi analizzano i movimenti di prezzo a breve termine e alcuni dei più noti e usati sono Moving Average Crossover, MACD (moving average convergence / divergence), RSI (relative strength index) e Bollinger Bands.

Analizzando i grafici passati mediante queste strategie è possibile ricavare i parametri per gli indicatori che meglio si adattano al mercato in analisi e riutilizzare l'indicatore, applicandolo ai dati futuri, per identificare nuovi trend.

1.4.1 Moving Average Crossover

Questa strategia prevede il calcolo di una o più medie mobili a diverso periodo, le cui intersezioni nel grafico definiscono un segnale di buy o sell.

Il moving average crossover si verifica quando, nel grafico di due medie mobili ciascuna basata su diversi periodi, le linee di queste medie mobili si incrociano. Non prevede la direzione futura ma mostra le tendenze. Questo indicatore utilizza due (o più) medie mobili, una più lenta e una più veloce. La più veloce è una media mobile a breve termine. Per i mercati azionari di fine giornata, ad esempio, può essere un periodo di 5, 10 o 25 giorni mentre la più lenta è la media mobile di medio o lungo termine (ad esempio periodo di 50, 100 o 200 giorni). Una media a breve termine è più veloce perché considera i prezzi solo per un breve periodo di tempo ed è quindi più reattiva alle variazioni giornaliere dei prezzi. D'altra parte, una media mobile a lungo termine è considerata più lenta in quanto incapsula i prezzi per un periodo più lungo, tuttavia tende ad attenuare i disturbi che si riflettono spesso nelle medie mobili a breve termine.

Una moving average, come una linea a sé stante, viene spesso sovrapposta nei grafici per indicare l'andamento dei prezzi. Un crossover si verifica quando una media mobile più veloce (cioè una media mobile di periodo più breve) attraversa una media mobile più lenta (cioè una media mobile di periodo più lungo).

In altre parole, accade quando la linea della media mobile del periodo più breve attraversa una linea della media mobile del periodo più lungo: questo punto di incontro viene utilizzato per comprare o vendere titoli (quando la media veloce supera quella lenta, si ha un segnale di *buy*, mentre *sell* se viceversa)



Figura: Andamento dei prezzi per il titolo TESLA/DOLLARO, accompagnato da due medie mobili a diverso periodo: fast_ma a 20 giorni mentre slow_ma a 100 giorni. Le frecce verdi e rosse indicano i segnali di buy e sell ricavati dalle intersezioni fra le medie

Ci sono diverse declinazioni della strategia a medie mobili, come ad esempio SMA (simple moving average), che utilizza come segnale di buy / sell il crossover fra i prezzi di chiusura e la media mobile semplice; EMA usa una media esponenziale; MACD (moving average convergence / divergence) usa come prima statistica la differenza fra la media veloce e quella lenta e, come seconda statistica, la media di queste differenze.

Calcolati gli indicatori statistici più adatti alla situazione e scelti i parametri migliori (come i periodi delle medie da usare), si è costruito un modello di trend del titolo e si è in grado di prevedere quando il trend si ripropone. Prendendo l'esempio di una moving average crossover con una media veloce e una lenta, analizzato che per la maggior parte dei casi si riscontra un guadagno vendendo quando la media lenta supera quella veloce e comprando quando accade l'opposto, allora si può affermare che sarà molto probabile guadagnare ancora, in futuro, ripetendo le medesime operazioni.

1.5 Ruolo delle intelligenze artificiali e scopo della tesi

L'apprendimento automatico e diverse tecniche hanno creato nuovi sistemi per individuare pattern, cosa che il cervello umano non è in grado di fare. Dal momento che la finanza è quantitativa, la AI nel trading azionario sta guadagnando terreno. Le società finanziarie hanno investito molto nell'intelligenza artificiale in passato e studiano e implementano le applicazioni finanziarie di machine learning e del deep learning nelle loro operazioni.

L'analisi tecnica è dunque implementata in sistemi automatici, che sono molto più efficienti di un trader umano e offrono diversi vantaggi. Le AI separano le informazioni prevedibili da qualsiasi "rumore casuale", gli algoritmi apprendono dall'accuratezza delle previsioni precedenti e si adeguano continuamente, abbastanza veloci da prevedere il mutamento della situazione del mercato anche in un orizzonte di pochi giorni. Possono imparare dai successi e fallimenti e riconfigurare ogni giorno le approssimazioni del funzionamento interno del mercato, poiché viene alimentato con nuovi dati. Oltre agli indiscutibili vantaggi computazionali di una macchina rispetto al cervello, ci sono altri aspetti da considerare che hanno portato alla crescita delle AI nel settore. Si può infondere la conoscenza degli esperti dell'ambito in un software che si auto adatta e migliora, senza avere gli svantaggi dell'errore umano; si può estendere per permettergli di gestire diversi ambiti, dal recupero della materia prima (i dati di trading) fino alla vendita del prodotto finito (vendita di strategie di investimento). È possibile costruire sistemi complessi che modellano intere realtà.

Sentymment è uno di questi software descritti. Si occupa di raccogliere dati, elaborarli e, attraverso una AI li analizza per produrre strategie di investimento. Il modulo di intelligenza artificiale è molto ampio: esistono molte AI che operano producendo costantemente nuovi segnali di buy / sell per le strategie e sono monitorate da una ulteriore, singola AI. Questo supervisore decide quale fra le molte AI sta avendo più successo e la promuove come attiva da utilizzare.

Lo scopo della tesi è sviluppare il supervisore e inserirlo all'interno del complesso sistema di Sentymment, impiegando tecniche di test per intelligenze artificiali che permettano di scegliere un rappresentante migliore fra una serie di agenti che operano indipendentemente.

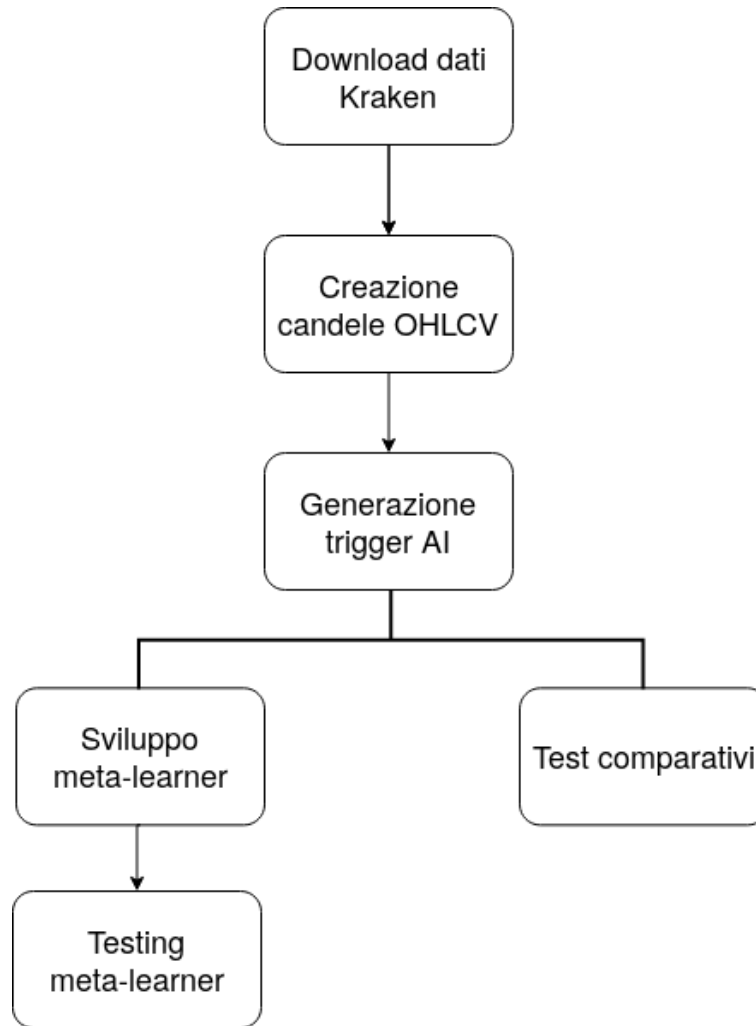


Figura: Diagramma delle attività svolte, spiegate nel dettaglio in ogni relativo capitolo. La fase preliminare di download dati comporta la successiva analisi esplorativa degli stessi e fornisce l'input per il resto del sistema. Il passo successivo è la creazione di candele OHLCV, cioè il tipo di dato utilizzato dalle AI. A questo punto la AI Sentiment può mettere in moto i suoi algoritmi per produrre le strategie di investimento, ovvero l'elenco di operazioni buy / sell (*trigger*) per ogni istante di tempo. Da qui si sono svolte due differenti attività: l'analisi dei risultati prodotti dalla AI (test comparativi) e lo sviluppo della AI supervisore (*meta-learner*), con successiva analisi e testing dei suoi risultati.

Capitolo 2

AI per financial prediction e Sentymment

La predizione del mercato azionario è stata una questione importante nel campo della finanza, ingegneria e matematica a causa del suo potenziale guadagno finanziario. Data la grande quantità di capitale negoziato attraverso il mercato azionario, questo è visto come una grande opportunità e ha richiamato molti investimenti. Inoltre, la previsione del mercato azionario comporta la sfida di dimostrare se il mercato finanziario è prevedibile o no. Dal momento che non c'è stato consenso sulla validità dell'ipotesi che nel mercato ci sia spazio per la predizione, i ricercatori hanno molto faticato per dimostrarne la prevedibilità[1]. L'*ipotesi del mercato efficiente* (EMH: Efficient Market Hypotesis) è un'ipotesi in economia finanziaria che afferma che i prezzi delle attività riflettono tutte le informazioni disponibili. Una conseguenza diretta è che è impossibile "battere il mercato" in modo coerente basandosi sul rischio poiché i prezzi di mercato dovrebbero reagire solo a nuove informazioni. È una teoria che fa sorgere molte controversie e come tale non ancora dimostrata, il che lascia spazio ai tentativi di sviluppo di un algoritmo che sia in grado di prevedere parte dell'andamento dei mercati.

Con l'avvento di computer più veloci e vaste informazioni su Internet, i mercati azionari sono diventati più accessibili agli investitori strategici o al pubblico generale. Internet fornisce la fonte principale di informazioni e ha un impatto significativo sui mercati azionari e le tecniche di estrazione e utilizzo delle informazioni a supporto del processo decisionale hanno raggiunto un'importanza critica. Per prevedere il mercato azionario con precisione, vari algoritmi e modelli di predizione sono stati proposti da molti ricercatori sia accademici che nell'industria.

2.1 Stato dell'arte delle AI

Da quando il mercato azionario è stato introdotto per la prima volta, molti hanno tentato di prevederne il comportamento utilizzando vari strumenti computazionali come *Linear egression* (LR), *Neural Network* (NNs), *Genetic Algorithm* (GA), *Support Vector Machine* (SVMs), *Case-based Reasoning* (CR) e altri.

Nell'ultimo decennio, Neural Network sono state ampiamente utilizzate e in molti casi hanno mostrato prestazioni migliori rispetto ad altri approcci. Una **rete neurale** (NN "Neural Network"), è un modello matematico/informatico di calcolo basato sulle reti neurali biologiche. Tale modello è costituito da un gruppo di interconnessioni di informazioni costituite da neuroni artificiali e processi che utilizzano un approccio di connessionismo di calcolo. Nella maggior parte dei casi una rete neurale artificiale è un sistema adattivo che cambia la propria struttura in base a informazioni esterne o interne che scorrono attraverso la rete stessa durante la fase di apprendimento.

Il primo modello di previsione del mercato azionario basato su Neural Network è stato implementato da White[2]. Ha usato *Feed Forward Neural Networks* (FFNN)¹ per decodificare regolarità precedentemente non rilevate nei movimenti di prezzo degli asset, come fluttuazioni di prezzo di azioni ordinarie, e ha mostrato come cercare tali regolarità usando FFNNs. Dal tentativo iniziale di White, un numero di ricercatori ha partecipato allo sviluppo di un modello di previsione del mercato azionario accurato. Phua et al. [3] ha utilizzato Neural Network con Genetic Algorithm per prevedere l'indice di scambio del mercato azionario di Singapore², ottenendo un tasso di accuratezza dell'81%. Kim e Han [4] hanno anche combinato Neural Network con Genetic Algorithm e pronosticato Korea Composite Stock Index 200³. He raggiunto l'82% di accuratezza nel prevedere sia le tendenze in aumento e in calo del mercato azionario.

Genetic Algorithm è stato studiato e dimostrato essere efficace nell'esplorare uno

¹Una rete neurale feed-forward ("rete neurale con flusso in avanti") o rete feed-forward è una rete neurale artificiale dove le connessioni tra le unità non formano cicli, differenziandosi dalle reti neurali ricorrenti. In questa rete neurale le informazioni si muovono solo in una direzione, avanti, rispetto a nodi d'ingresso, attraverso nodi nascosti (se esistenti) fino ai nodi d'uscita. Nella rete non ci sono cicli. Le reti feed-forward non hanno memoria di input avvenuti a tempi precedenti, per cui l'output è determinato solamente dall'attuale input.

²L'indice FTSE Straits Times (STI) è un indice di borsa ponderato per la capitalizzazione che è considerato l'indice di riferimento per il mercato azionario di Singapore. Tiene traccia delle prestazioni delle prime 30 società quotate alla Borsa di Singapore.

³Korea Composite Stock Index (KOSPI): indice di borsa azionario della Corea del Sud, che ospita più di 780 società quali Samsung, Hyundai e Kia motors

spazio complesso in modo adattivo, guidato dal meccanismo evolutivo di riproduzione, crossover e mutazione.

Un **algoritmo genetico** è un algoritmo euristico⁴ utilizzato per tentare di risolvere problemi di ottimizzazione per i quali non si conoscono altri algoritmi efficienti di complessità lineare o polinomiale. L'aggettivo "genetico", ispirato al principio della selezione naturale ed evoluzione biologica, deriva dal fatto che gli algoritmi genetici attuano dei meccanismi concettualmente simili a quelli dei processi biochimici scoperti dalla biologia.

In sintesi gli algoritmi genetici consistono in algoritmi che permettono di valutare diverse soluzioni di partenza (come se fossero diversi individui biologici) e che ricombinandole ed introducendo elementi di disordine (mutazioni genetiche casuali) producono nuove soluzioni che vengono valutate scegliendo le migliori nel tentativo di convergere verso soluzioni "di ottimo". Nonostante questo utilizzo nell'ambito dell'ottimizzazione, data la natura intrinsecamente casuale dell'algoritmo genetico, non vi è modo di sapere a priori se sarà effettivamente in grado di trovare una soluzione accettabile al problema considerato. La soluzione trovata sarà frutto di una procedura casuale e quindi difficilmente riproducibile.

Di recente, molti ricercatori hanno impiegato Genetic Algorithm per migliorare le prestazioni delle tecniche di AI. Nell'ambito delle Neural Network, Genetic Algorithm è comunemente usato per selezionare la topologia di rete e determinare il numero ottimale di layer nascosti e di elementi di elaborazione.

Molti ricercatori hanno confrontato le Neural Network con approcci statistici per il riconoscimento di pattern [5][6]. Yoon et al. [7] ha sottolineato la superiorità di Neural Network sull'analisi discriminante classica. Garliauskas [5] ha inoltre esaminato la previsione del mercato azionario utilizzando Neural Network in combinazione con un approccio alle *funzioni kernel*⁵ e il metodo dell'errore di previsione ricorsivo⁶. Garliauskas [5] ha concluso che, nel prevedere le serie temporali finanziarie, le Neural

⁴Un algoritmo euristico, o *euristica*, è un particolare tipo di algoritmo progettato per risolvere un problema più velocemente, nel caso in cui i metodi classici siano troppo lenti nel calcolo (ad esempio, in caso di elevata complessità computazionale) o per trovare una soluzione approssimata, nel caso in cui i metodi classici falliscano nel trovare una soluzione esatta.

⁵Le funzioni kernel sono una classe di algoritmi per l'analisi di pattern, schemi, il cui elemento maggiormente conosciuto sono le Support Vector Machine (SVM). Lo scopo generale dell'analisi di pattern è di trovare e studiare tipi generici di relazioni. Le funzioni kernel si avvicinano al problema mappando i dati in uno spazio di caratteristiche multidimensionale, trasformando i dati in un insieme di punti dello spazio euclideo.

⁶Errore di predizione: è una misura di quanto bene il modello prevede la variabile di risposta, o di come i campioni sono classificati nella categoria corretta.

Network hanno prestazioni migliori rispetto ai metodi statistici classici. In generale, numerosi studi hanno dimostrato che le Neural Network hanno la capacità di prevedere i mercati azionari in modo più accurato rispetto ad altri metodi [7][8][9][10].

2.1.1 Supporto di informazioni relative agli eventi

Sebbene gli studi precedenti fossero limitati ad analisi quantitativa, diversi studi recenti sono stati eseguiti sulla base dell'analisi qualitativa. Molti credevano che l'integrazione di una conoscenza degli eventi nelle Neural Network risultasse promettente per una migliore previsione del mercato azionario [12][13][14][15]. Kohara [14] ha studiato alcune metodologie per migliorare i modelli predittivi multivariati⁷ nella previsione del prezzo delle azioni utilizzando Neural Network e conoscenza pregressa. Ha usato le informazioni dai titoli dei giornali per migliorare la capacità di previsione. I loro risultati sperimentali hanno dimostrato che l'utilizzo della conoscenza degli eventi unita alle Neural Network ha ridotto significativamente il tasso di errore di previsione intorno al livello di significatività del 5% con profitto del 40%.

Inoltre, Hong e Han [12] hanno confrontato le Neural Network con informazioni sugli eventi rispetto al modello *Random Walk* (RW)⁸ e rispetto a Neural Network senza informazioni sull'evento. Neural Network con informazioni sull'evento hanno una media dello 0,527% di errore, che è un errore minore rispetto ad altri modelli. Hanno dimostrato che le previsioni basate su Neural Network sono notevolmente superiori a Random Walk e che l'effetto dell'informazione sugli eventi esiste.

Inoltre, nella previsione mercato azionario, molti ricercatori hanno riconosciuto che fattori qualitativi, quali effetti politici ed eventi internazionali, hanno svolto un ruolo molto importante. Hanno dunque proposto che un sistema di previsione del mercato azionario debba incorporare fattori sia quantitativi che qualitativi. I loro studi hanno infine dimostrato che le Neural Network che si basano sui fattori sia quantitativi che qualitativi sono di gran lunga superiori a quelle basate solo su fattori quantitativi [16][17].

⁷Il modello multivariato è un comune strumento statistico che utilizza più variabili per prevedere i possibili risultati. Gli analisti della ricerca utilizzano modelli multivariati per prevedere i risultati degli investimenti in diversi scenari al fine di comprendere l'esposizione che un portafoglio ha a particolari rischi. Ciò consente ai gestori di portafoglio di mitigare meglio i rischi identificati attraverso l'analisi della modellazione multivariata. La **simulazione Monte Carlo** è un modello multivariato ampiamente utilizzato che crea una distribuzione di probabilità che aiuta a definire una gamma di possibili risultati di investimento.

⁸In matematica, una passeggiata aleatoria (Random Walk) è la formalizzazione dell'idea di prendere passi successivi in direzioni casuali. Matematicamente parlando, è il processo stocastico più semplice.

2.1.2 Punti chiave nelle predizioni finanziarie

Numerosi ricercatori hanno mostrato il loro ufficiale punto di vista sull'*ipotesi del mercato efficiente* (EMH: efficient market Hypotesis) in ambito accademico [18][19]. EMH afferma che l'attuale prezzo di mercato riflette l'assimilazione di tutte le informazioni disponibili[20]. L'informazione rilevante per un mercato è contenuta nei prezzi e ogni volta che sorgono nuove informazioni, il mercato corregge stesso e le assorbe, quindi, il mercato è efficiente e non c'è spazio per la predizione[21].

È in corso un lungo dibattito sulla validità di EMH, ma non è stato raggiunto un consenso. Negli ultimi anni, molti i ricercatori hanno affermato che sicuramente EMH deve essere falso[21]. Diversi studi sono stati condotti sui dati dei mercati azionari al fine di dimostrare che il mercato è prevedibile. Se qualche sistema di calcolo avesse la capacità di mostrare previsioni con ragionevole accuratezza, basandosi sui dati storici di mercato, allora la validità di EMH potrebbe essere messa in discussione.

Tsibouris e Zeidenberg[22] hanno utilizzato le Neural Network per prevedere il mercato azionario basandosi soltanto sui prezzi passati come input. I loro risultati empirici hanno mostrato un certo livello di capacità predittiva e respinto la forma debole di EMH⁹.

Inoltre, Eyden [23] ha sviluppato un sistema che modella le prestazioni della borsa azionaria di Johannesburg e il loro sistema ha fornito prove significative dimostrando la sua capacità di prevedere le direzioni del mercato azionario, quindi, di nuovo, dissentendo da EMH.

Dalla parte opposta, invece, vari studi sono basati sull'ipotesi di mercato efficiente. Fung et al. [13] ha indagato sull'immediato impatto delle notizie di giornale sulle serie storiche basate su EMH. Hanno presentato un sistema basato su EMH per prevedere il comportamento futuro del mercato azionario utilizzando informazioni non quantificabili (notizie dai giornali) e le previsioni sono state fatte solo in base al contenuto dell'articolo. Alla fine il loro risultato empirico ha indicato che l'approccio basato su EMH è più redditizio rispetto alla semplice strategia di trading basata su Buy-Hold¹⁰. Come è stato discusso, diversi studi hanno concluso di accettare o confutare EMH. La questione dell'efficienza del mercato dunque non è ancora pienamente

⁹La forma debole dell'*ipotesi di mercato efficiente* (EMH, efficient market hypothesis) afferma che i prezzi passati, i valori storici e le tendenze non possono prevedere i prezzi futuri. I prezzi delle azioni riflettono soltanto tutte le informazioni attuali

¹⁰Buy and Hold: strategia di investimento, usata spesso come base per confrontare altre strategie. Un investitore acquista titoli e li detiene a lungo, con l'obiettivo di aumentare gradualmente col tempo il loro valore. Il guadagno segue l'andamento del mercato: il valore del titolo può salire o scendere. È la strategia minimale da "battere" se si vogliono sviluppare nuove strategie: compra e detieni, senza fare niente.

indagata e porta alla necessità di ulteriori ricerche.

Un certo numero di ricercatori ha usato la cronologia dei dati numerici delle serie temporali per la previsione dei mercati azionari e ha raggiunto una ragionevole accuratezza delle previsioni rifiutando l'ipotesi di EMH [23][22].

Tuttavia, ci sono vari fattori che influenzano i prezzi delle azioni, come le prestazioni e la robustezza di un'azienda, le tendenze del mercato, la psicologia degli investitori, il coinvolgimento del governo, i cambiamenti nell'attività economica e così via. Quindi, molti ricercatori hanno concordato sull'esistenza di una correlazione significativa tra gli eventi che rappresentano i fattori sopra citati e i mercati azionari [13][14]. Hanno utilizzato le informazioni relative agli eventi in aggiunta a quelle numeriche delle serie temporali e raggiunto un certo livello di precisione nelle previsioni. Poiché, però, ogni evento ha un diverso livello di impatto in termini di intervallo di tempo in cui l'evento stesso può influenzare il mercato, ogni evento dovrebbe essere ponderato in base al suo livello di impatto, che è definito sulla base della conoscenza preliminare costruita in precedenza. In aggiunta, il web è visto come la principale fonte di informazione sugli eventi per la previsione di mercato, contenendo le più aggiornate e preziose informazioni. Pertanto, sono necessarie tecniche automatiche di *web mining*¹¹ per ottenere un livello più elevato di accuratezza della previsione in tempi brevi.

¹¹Il web mining è la raccolta sistematica di dati provenienti dal web, organizzati e visualizzati al fine di far emergere informazioni utili. Prevede l'impiego di tecniche di data mining al fine di estrapolare dati da server e pagine web.

2.1.3 Metodi di previsione

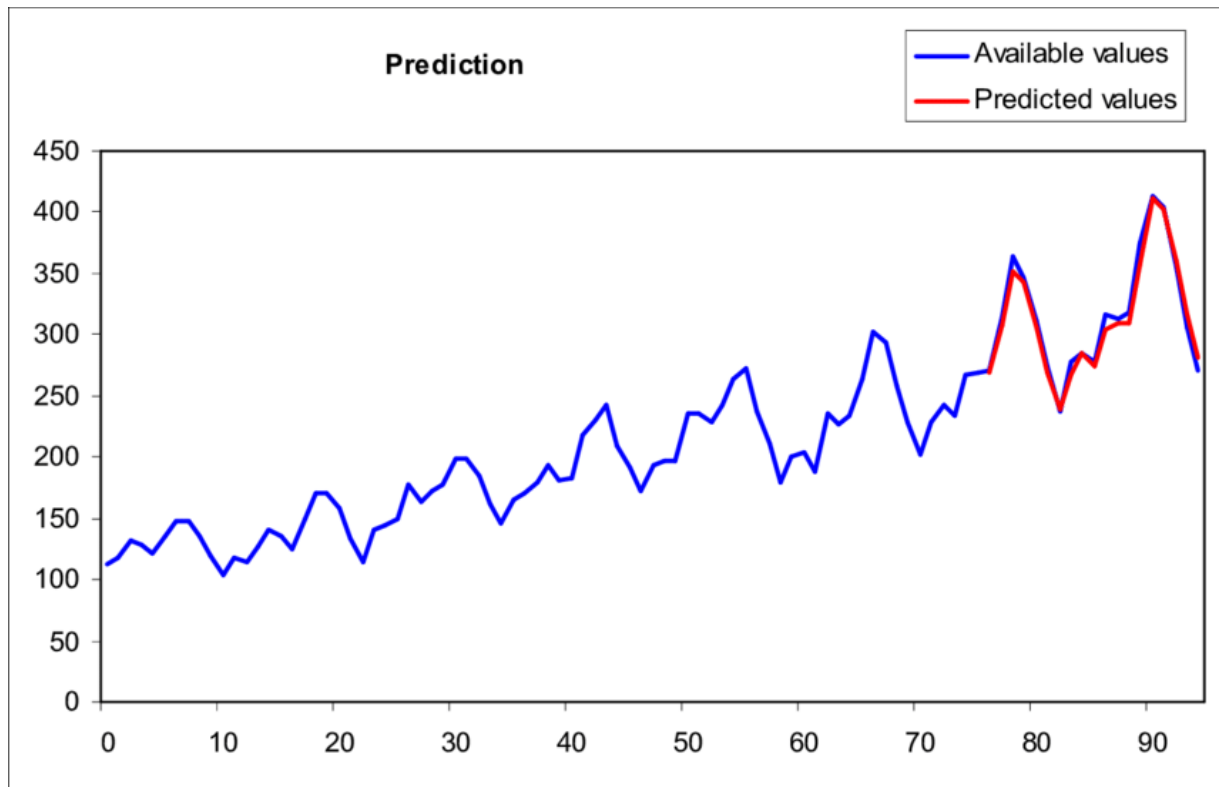


Figura: esempio di utilizzo di una tecnica per predizione. Si suppone di analizzare il grafico dell'andamento di un titolo (serie storica), in cui le ascisse sono il tempo e le ordinate il prezzo dell'asset. Come si può notare, molto spesso nelle serie storiche finanziarie, i pattern si ripetono nel tempo e applicando metodi di regressione o intelligenza artificiale è possibile prevedere quando questi si ripresenteranno.

Predizione tradizionale di serie storiche

I modelli statistici tradizionali sono ampiamente utilizzati in economia per la previsione di serie storiche¹². Questi modelli sono in grado di modellare relazioni lineari tra fattori che influenzano il mercato e il valore del mercato. In economia, ci sono due tipi di previsione delle serie: univariata (regressione semplice) e multivariata (regressione multivariata) [19]. Un esempio ampiamente usato di modello univariato è

¹²In statistica descrittiva, una serie storica o temporale ("time series") si definisce come un insieme di variabili casuali ordinate rispetto al tempo, ed esprime la dinamica di un certo fenomeno nel tempo. Le serie storiche vengono studiate sia per interpretare un fenomeno, individuando componenti di trend, di ciclicità, di stagionalità e/o di accidentalità, sia per prevedere il suo andamento futuro.

Box-Jenkins¹³, che contiene solo una variabile nell'equazione ricorrente. Mostra un processo complesso di adattamento dei dati ai parametri appropriati del modello. Nelle loro equazioni sono inclusi i valori passati delle medie mobili dei prezzi.

Sebbene Box-Jenkins mostri una buona capacità di previsione a breve termine, richiede una grande quantità di dati per fornire un'alta precisione [24]. I modelli multivariati sono modelli univariati ampliati per "scoprire i fattori causali che incidono sul comportamento dei dati" [25]. Come implica il nome, l'equazione del modello contiene più di una variabile. Un certo numero di studi ha confrontato il modello statistico multivariato con Neural Network. Sebbene i modelli multivariati siano stati ampiamente utilizzati per prevedere i mercati azionari, diverse tecniche di machine learning stanno ora sostituendo i loro ruoli.

In letteratura, molti ricercatori hanno affermato che le Neural Network superano notevolmente i metodi statistici tradizionali [5][6]. Lawrence [1] ha confrontato le Neural Network con tecniche statistiche e di regressione. Ha usato un sistema che utilizza Neural Network con Genetic Algorithm. Il sistema con le AI ha mostrato la sua capacità di prevedere correttamente i movimenti del mercato il 92% delle volte, mentre Box-Jenkins ha mostrato un tasso di precisione soltanto del 60%. Inoltre, diversi sistemi che utilizzano Neural Network hanno mostrato prestazioni costantemente migliori rispetto a modelli di regressione lineare multipla [26] [27]. Una delle ricerche più peculiari è stata eseguita da Yoon et al. [7]. Il loro sistema basato su Neural Network ha predetto la tendenza del prezzo delle azioni con il 91% di precisione, rispetto al 74% prodotto utilizzando l'*analisi discriminante multipla*¹⁴. Così, secondo loro, si può concludere che Neural Network hanno costantemente prestazioni migliori rispetto alle tecniche statistiche e di regressione.

¹³Il modello autoregressivo a media mobile, detto anche ARMA, è un tipo di modello matematico lineare che fornisce istante per istante un valore di uscita basandosi sui precedenti valori in entrata e in uscita. A volte denominato modello di Box-Jenkins dal nome dei suoi inventori, viene utilizzato in statistica per lo studio delle serie storiche dei dati.

¹⁴Multiple Discriminant Analysis (MDA) è una tecnica statistica di riduzione della dimensionalità multivariata. È utilizzata dai pianificatori finanziari per valutare i potenziali investimenti quando devono essere prese in considerazione diverse variabili. Questa tecnica riduce le differenze tra alcune variabili in modo che possano essere confrontate con un'altra variabile. L'analisi discriminante multipla è correlata all'analisi discriminante, che aiuta a classificare un set di dati impostando una regola o selezionando un valore che fornirà la separazione più significativa.

Neural Network

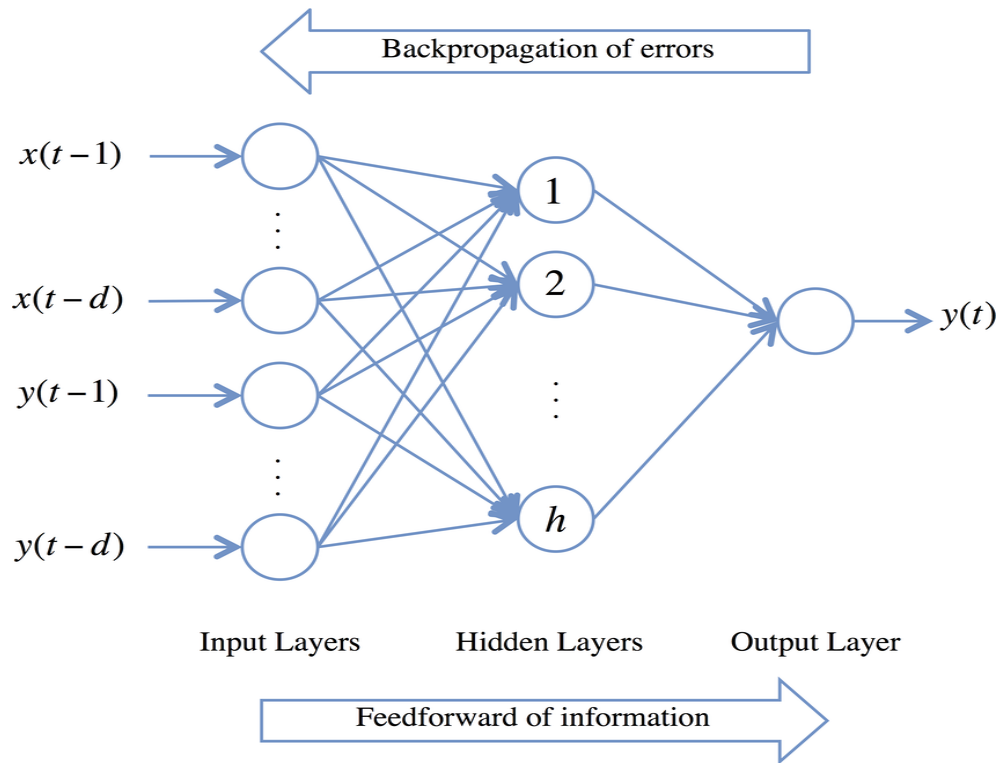


Figura: Schema semplificato di Backpropagation Neural Network. Una Neural Network è un'interconnessione di un gruppo di nodi chiamati neuroni. Ha un livello di neuroni di input, un numero variabile di livelli intermedi (*hidden layers*) e un livello di output. I segnali si propagano dagli input attraverso gli strati nascosti, fino all'output. Nel caso di backpropagation, durante il training i pesi e gli altri parametri della rete sono modificati propagando l'errore dall'output verso gli input.

In letteratura è stato dimostrato che Neural Network offre la capacità di prevedere le direzioni del mercato in modo più preciso rispetto ad altre tecniche esistenti. La capacità delle Neural Network di scoprire relazioni non lineari tra le coppie input / output di training li rendono ideali per la modellazione non sistemi dinamici non lineari come i mercati azionari [1]. Uno degli vantaggi è la capacità di poter apprendere le relazioni attraverso i dati stessi anziché considerare la forma funzionale della relazione. Dato che le Neural Network sono conosciute come approssimatori universali, ogni relazione può essere modellata con un qualsiasi grado di precisione, quando sufficienti dati vengono forniti per la modellazione. Forniscono inoltre un livello di tolleranza al rumore e alla rappresentazione incompleta dei dati.

Un altro vantaggio è che Neural Network hanno proprietà di apprendimento adattive non lineari, non parametriche e hanno un effetto molto pratico nella modellazione e predizione. La loro natura non lineare mostra un grande potenziale per risolvere molte problemi complessi. A causa delle caratteristiche di cui sopra dei mercati azionari, le Neural Network possono essere applicate alla predizione del mercato azionario. In primo luogo, è difficile modellare i dati di borsa a causa della loro complessità, quindi, il modello non lineare è vantaggioso. Inoltre, è spesso richiesto un ampio set di serie temporali correlate in input per modellare uno specifico asset.

Parlando invece degli aspetti negativi, Neural Network ha il problema della scatola nera, cioè che non rivela il senso e il valore di ogni variabile e il modo in cui pesa le indipendenti variabili [1]. Dato che il ruolo individuale di ogni variabile non può essere determinato, è impossibile da capire come la rete produce il prezzo futuro delle azioni.

Un altro grosso problema con le Neural Network è il sovrallenamento. Quando le reti si adattano troppo bene ai dati, il sistema perde la capacità di generalizzare. Siccome la capacità di generalizzazione è fondamentale per prevedere lo stock futuro prezzi, il sovrallenamento è un problema serio. Questo di solito avviene per due motivi principali, quando le Neural Network hanno troppi nodi o periodi di training troppo lunghi (*epoch*¹⁵). Tuttavia, il sovrallenamento può essere prevenuto eseguendo procedure di test e train o *cross validation*.

Per quanto riguarda il rumore tremendo e le caratteristiche non stazionarie nei dati di borsa, Lawrence et al. [1] ha sottolineato che, quando l'addestramento delle Neural Network è difficile per dati molto rumorosi, le reti cadono in una soluzione ingenua come prevedere sempre l'output il più comune. Inoltre, le Neural Network hanno alcune limitazioni nell'apprendimento dei pattern quando i dati di input hanno un'alta dimensionalità. Dash e Liu [28] pongono l'accento sulla selezione delle *feature*¹⁶ e suggeriscono che ridurre il numero di variabili in input a volte porta a un miglioramento di prestazioni del modello per un determinato set di dati. La riduzione e trasformazione delle feature irrilevanti o ridondanti possono ridurre il tempo di esecuzione e produrre risultati più generalizzati [28].

¹⁵Un epoch è quando un *intero* dataset di training viene passato in avanti e indietro attraverso la rete neurale solo una volta per allenarla.

¹⁶Nell'apprendimento automatico e nel riconoscimento di pattern, una "feature" (caratteristica) è una proprietà misurabile individuale o caratteristica di un fenomeno osservato. La scelta di feature informative, discriminanti e indipendenti è un passaggio cruciale per algoritmi efficaci nel riconoscimento, nella classificazione e nella regressione dei modelli. Il concetto di feature è correlato a quello della variabile esplicativa utilizzata nelle tecniche statistiche come la regressione lineare.

Ricerche recenti tendono a utilizzare diverse tecniche di AI *ibride*¹⁷. Hiemstra [29] ha proposto *sistemi esperti fuzzy*¹⁸ per prevedere i rendimenti del mercato azionario. Suggerì che combinando le Neural Network con la logica fuzzy si possa catturare la complessità del mapping funzionale e non è richiesta la specifica della funzione da approssimare. Tsaih et al. [30] ha integrato la tecnica basata su regole e le Neural Network per prevedere la direzione di cambio dello stock S&P 500¹⁹ *futures*²⁰ su base giornaliera.

Support Vector Machine

Support Vector Machine (SVM) basato su teoria dell'apprendimento statistico, è stato sviluppato da Vapnik [31] e i suoi colleghi alla fine degli anni '70. È diventato un argomento caldo di studio intensivo grazie al suo successo applicato a compiti di classificazione e regressione, soprattutto nella previsione di serie temporali e applicazioni correlate alla finanza [32]. Support Vector Machine è una tipologia molto specifica di algoritmi di apprendimento caratterizzato dal controllo della capacità della funzione di decisione, l'uso delle funzioni del kernel e la dispersione della soluzione [33]. Fondato sulla teoria del *principio di minimizzazione del rischio strutturale*²¹, in cui stima una funzione minimizzando l'errore di generalizzazione, Support Vector Machine si dimostra alla fine molto resistente al problema del sovrallenamento (*overtraining*), raggiungendo un ottimo risultato nella generalizzazione.

¹⁷Un sistema intelligente *ibrido* indica un sistema software che impiega, in parallelo, una combinazione di metodi e tecniche provenienti da diversi sottocampi di intelligenza artificiale, come, ad esempio: *neuro-fuzzy systems*, *evolutionary neural network*, *genetic fuzzy systems*, *reinforcement learning* con *fuzzy*, *neural*, o *evolutionary*, etc.

¹⁸La logica fuzzy è una logica in cui si può attribuire a ciascuna proposizione un grado di verità diverso da 0 e 1 e compreso tra di loro. È una logica polivalente, ossia un'estensione della logica booleana. Con grado di verità o valore di appartenenza si intende quanto è vera una proprietà: questa può essere, oltre che vera (= a valore 1) o falsa (= a valore 0) come nella logica classica, anche parzialmente vera e parzialmente falsa.

¹⁹L'indice Standard & Poor 500, noto come S&P 500 o semplicemente S&P, è stato realizzato da Standard & Poor's nel 1957 e segue l'andamento di un paniere azionario formato dalle 500 aziende statunitensi a maggiore capitalizzazione.

²⁰I futures (future contracts) sono contratti a termine standardizzati per poter essere negoziati facilmente in una borsa valori, e sanciscono l'impegno a un acquisto differito a un prezzo prefissato.

²¹La minimizzazione del rischio strutturale (SRM) è un principio induttivo in uso nell'apprendimento automatico. Comunemente, in machine learning, un modello generalizzato deve essere selezionato a partire da un set di dati finito, con il conseguente problema di overfitting: il modello diventa troppo fortemente adattato alle particolarità dei dati di training e generalizza male i nuovi dati. Il principio SRM affronta questo problema bilanciando la complessità del modello con il suo successo nell'adattare i dati di addestramento.

Un'altra proprietà chiave è che addestrare Support Vector Machine equivale a risolvere un problema di programmazione linearmente vincolato. Pertanto, la soluzione di Support Vector Machine è relativamente unica e globalmente ottimale, a differenza del training delle Neural Network, che richiede un'ottimizzazione non lineare con pericolo di rimanere bloccati ai minimi locali.

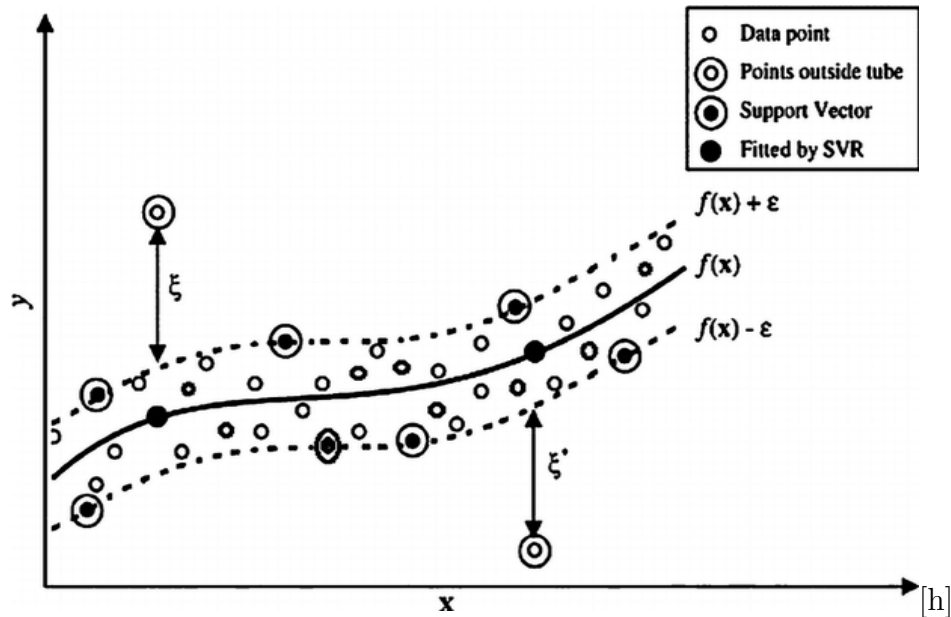


Figura: Support Vector Machine applicato al compito di regressione: nel caso di funzioni non lineari, queste vengono mappate in spazi multidimensionali per mezzo di metodi kernel. Dopodichè viene applicata la tecnica dei vettori di supporto per delineare gli estremi della funzione da approssimare. Si ottiene un'approssimazione che, in ogni punto, discosta al massimo di ε dai valori noti della funzione.

Di recente, diverse applicazioni di Support Vector Machine a problemi di previsioni finanziarie sono state documentate [34][35][36]. Uno degli studi ben noti che utilizzano Support Vector Machine per la predizione del mercato è stato eseguito da Kim [34]. Ha applicato Support Vector Machine alle previsioni finanziarie e confrontato il risultato rispetto a *backpropagation*²² Neural Network e Case Based Reasoning

²²In machine learning, in particolare deep learning, backpropagation è un algoritmo ampiamente utilizzato nel training di reti neurali feedforward per l'apprendimento supervisionato. L'algoritmo di backpropagation funziona calcolando il gradiente della funzione di perdita rispetto a ciascun peso della *chain rule*, calcolando il gradiente uno strato alla volta, ripetendo all'indietro dall'ultimo strato per evitare calcoli ridondanti di termini intermedi.

(CBR). I risultati sperimentali hanno mostrato che Support Vector Machine ha avuto performance migliori di Neural Network e Case Based Reasoning. Poiché Support Vector Machine si basa sul principio di minimizzazione del rischio strutturale, la generalizzazione implementata risulta migliore rispetto alle tecniche tradizionali.

Case Based Reasoning

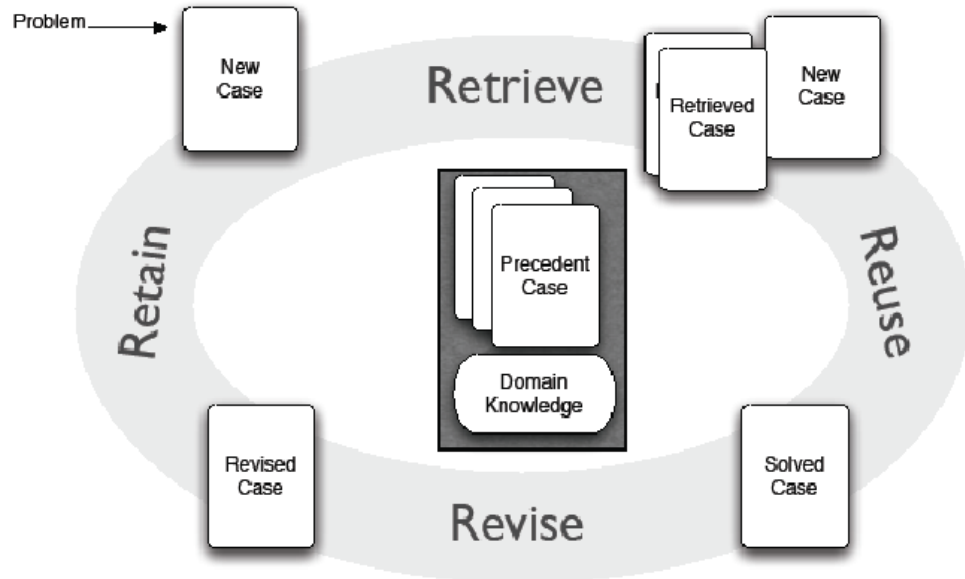


Figura: Il processo di risoluzione di nuovi problemi basandosi sulle soluzioni di problemi anteriori, esempio generale di algoritmo CBR. Dato un problema, recuperare in memoria dei casi rilevanti per risolverlo (*Retrieve*). Adattare la soluzione del caso precedente al problema attuale (*Reuse*). Avendo mappato la soluzione precedente al caso attuale, bisogna provare la nuova soluzione e, se necessario, rivedere la nuova soluzione (*Revise*). Dopo che la soluzione è stata adattata al problema attuale, memorizzare l'esperienza come nuovo caso (*Retain*).

Sebbene le Neural Network offrano capacità di apprendimento relativamente buone rispetto ad altre tecniche, non possono sempre esplicitare il perché arrivano a una particolare soluzione. Inoltre non garantiscono sempre una soluzione completamente certa, non assicurano di arrivare ripetutamente alla stessa soluzione dato lo stesso dataset di training, né possono garantire di trovare la soluzione ottimale[37]. Tuttavia,

a differenza delle reti neurali, in genere i *sistemi esperti*²³ forniscono spiegazioni per le loro soluzioni. I sistemi esperti acquisiscono principalmente la conoscenza da individui esperti nell'ambito. Le organizzazioni hanno conoscenze collettive e competenze che hanno accumulato nel corso degli anni. Queste conoscenze possono essere acquisite e archiviate utilizzando Case Based Reasoning. È una tecnica di ragionamento che riutilizza i casi passati per trovare una soluzione al nuovo problema. Case Based Reasoning non solo acquisisce la conoscenza dell'organizzazione ma fornisce anche spiegazioni per le soluzioni derivate. Per questa ragione è comunemente applicato a molti problemi.

Kim [38] ha proposto un nuovo modello ibrido di Genetic Algorithm e Case Based Reasoning per la previsione del mercato azionario. Dal suo studio preliminare ha scoperto che la scelta del peso delle feature e del sottoinsieme delle stesse è molto importante per migliorare le prestazioni delle predizioni del sistema. Ha quindi usato Genetic Algorithm come metodo di selezione del sottoinsieme di feature nel sistema Case Base Reasoning. Ha confrontato i risultati di quattro diversi modelli per testare l'efficacia del modello proposto. I modelli a confronto sono: CBR convenzionali (COCBR), pesatura delle feature utilizzando Genetic Algorithm per CBR (FWCBR), selezione delle feature usando Genetic Algorithm per CBR (FSCBR) e ottimizzazione simultanea usando Genetic Algorithm per CBR (SOCBR). I risultati empirici hanno mostrato che SOCBR raggiunge una precisione di previsione superiore rispetto a COCBR, FWCBR e FSCBR. Kim ha concluso che il modello ibrido di GA e CBR offre un valido approccio alternativo alla previsione del mercato azionario.

Informazione sugli eventi

Nei mercati azionari, ci sono molti fattori che possono influenzare il prezzo dei titoli. Questi fattori possono essere derivati dai comunicati stampa sulle piccole aziende o dalle notizie delle superpotenze dell'economia mondiale. Questi episodi sono chiamati *events* [39]. Il motivo principale per cui incorporare conoscenza degli eventi nella previsione del mercato azionario si basa sul presupposto che il prezzo futuro di un'azione dipende in parte da vari aspetti politici ed eventi internazionali, oltre ai vari indicatori economici. Pertanto, molti studi hanno utilizzato informazioni sugli eventi (fattori qualitativi) nonché dati quantitativi nella previsione dei mercati azionari.

Uno degli studi popolari che utilizzano la conoscenza degli eventi e la *conoscenza*

²³Un sistema esperto è un programma che cerca di riprodurre le prestazioni di una o più persone esperte in un determinato campo di attività, ed è un'applicazione o una branca dell'intelligenza artificiale. Sono spesso implementati attraverso euristiche o fuzzy logic e si dividono in due categorie principali: basati su regole e basati su alberi.

pregressa è stato realizzato da Kohara et al. [14]. I ricercatori hanno incorporato la conoscenza preliminare nella previsione di mercato, come informazioni sui giornali su eventi nazionali ed esteri. La conoscenza dell'evento è estratta dai titoli dei giornali in conformità con alcune conoscenze preliminari. La conoscenza pregressa è l'informazione che deriva da precedenti esperienze, pertanto, in base alla conoscenza precedente, decisioni differenti possono essere prese se un evento particolare può influenzare positivamente o meno le tendenze del mercato azionario.

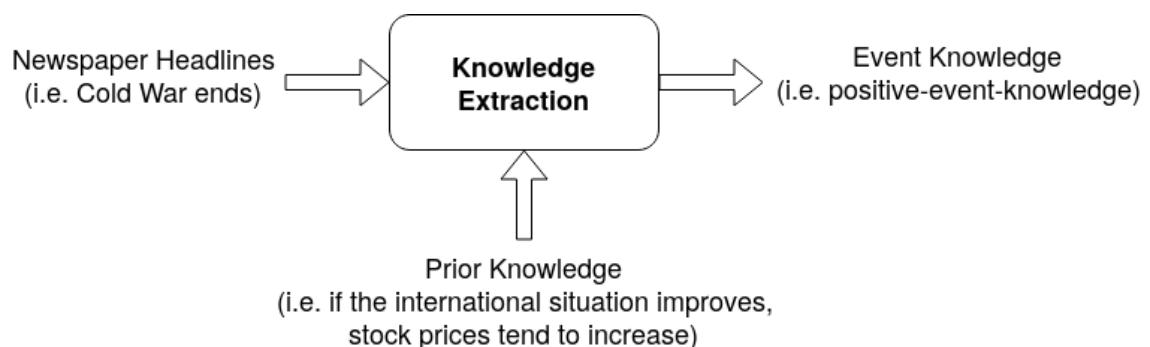


Figura: Estrazione di conoscenza sugli eventi: basandosi sulle notizie di giornale e avendo coscienza dell'impatto che possono avere certi eventi sull'economia, è possibile stimare se un certo evento può influire positivamente o negativamente sul mercato.

Kohara et al. [14] hanno selezionato diversi indicatori economici (tasso di interesse, prezzo del petrolio grezzo e media dei prezzi di chiusura di New York Dow Jones) e forniti insieme alla conoscenza degli eventi come dati di training per Neural Network. I loro risultati sperimentali hanno mostrato che incorporare la conoscenza di eventi ha migliorato la capacità di previsione delle Neural Network, riducendo il tasso di errore del 5% del livello di significatività. Intanto, risulta cruciale *come* incorporare l'impatto delle informazioni sulle notizie nei modelli di serie storiche. Maheu e McCurdy [40] hanno specificato un modello GARCH²⁴ con salti per serie di ritorno, che può essere direttamente misurato dai dati dei prezzi. Si è supposto che il processo latente delle notizie abbia due componenti separati, eventi di cronaca normali e insoliti. Queste differenti notizie sono identificate attraverso il loro impatto sulla *volatilità* del rendimento²⁵.

²⁴In econometria, un modello autoregressivo a eteroschedasticità condizionata generalizzato o modello GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) è un modello utilizzato nell'analisi delle serie storiche. È una funzione dei valori assunti dal processo agli istanti precedenti.

²⁵In finanza, la volatilità è una misura della variazione percentuale del prezzo di uno strumento finanziario nel corso del tempo. Corrisponde alla *Deviazione standard*. Di solito indica il livello di rischio associato ad un titolo: possedendo un titolo più volatile si è più esposti alle oscillazioni del

I modelli precedenti forniscono quadri generali per incorporare gli impatti delle notizie di giornale. Questi metodi non forniscono però un approccio per la distinzione delle notizie influenti o significative per un determinato titolo, a fronte di migliaia di articoli di ogni genere. Pertanto questi metodi, nella pratica, non possono garantire un miglioramento significativo.

In letteratura, un certo numero di ricercatori ha affermato che i prezzi delle azioni sono significativamente correlati con le informazioni relative agli eventi e molti hanno tentato di utilizzare sia informazioni sugli eventi che dati numerici di serie temporali come dati in input. Fawcett e Provost [41] hanno formulato il problema di previsione del mercato come un'attività di monitoraggio della relazione tra le notizie e i prezzi delle azioni. Fung et al. [13] hanno proposto un sistema che prevede il cambiamenti nell'andamento del mercato analizzando l'influenza di informazioni qualitative (notizie di giornale). In particolare, hanno studiato l'impatto immediato delle notizie sulle serie storiche basate sull'ipotesi di mercato efficiente. Hanno proposto un nuovo algoritmo statistico di *segmented regression*²⁶ per identificare le tendenze nelle serie temporali. In letteratura viene sempre più riconosciuta l'incorporazione della conoscenza pregressa e conoscenza degli eventi nei modelli di previsione come Neural Network.

Informazione sugli eventi nel Web

L'evoluzione di Internet con l'infrastruttura di informazioni globale ha provocato un'esplosione della quantità di informazioni disponibili. Enormi moli di dati relativi a eventi che possono avere grande influenza sul mercato azionario sono disponibile sul web. Considerando che i giornali vengono aggiornati una o due volte al giorno, le fonti di notizie in tempo reale sono aggiornate frequentemente quasi istantaneamente. Queste informazioni non contengono solo notizie globali e regionali ma anche citazioni preziose di influenti banchieri, politici e analisti finanziari. Le informazioni che consistono in notizie o citazioni sono viste come principali motori del mercato delle obbligazioni, delle valute e azionario nel mondo. Hong e Han [42] hanno affermato che, con l'aumento della popolarità di Internet, molti giornali espanderanno i loro servizi fornendo notizie sul web, per essere più competitivi e aumentare il profitto.

suo valore.

²⁶La regressione segmentata (segmented regression), nota anche come piecewise regression o broken-stick regression, è un metodo di analisi di regressione in cui la variabile indipendente è suddivisa in intervalli e un segmento di linea separato è adattato a ciascun intervallo. L'analisi di regressione segmentata può anche essere eseguita su dati multivariati partizionando le varie variabili indipendenti. La regressione segmentata è utile quando le variabili indipendenti, raggruppate in gruppi diversi, mostrano relazioni diverse tra le variabili in queste regioni.

Le notizie includono articoli sulla situazione politica, le condizioni sociali, eventi internazionali, politiche governative, psicologia dei trader e così via; informazioni che vediamo e comprendiamo attraverso Internet. Tali informazioni sono formulate sotto forma di testo, indicati come documenti, quindi tecniche di *text mining*²⁷ sono necessarie.

Hong e Han [42] hanno introdotto un sistema automatizzato (KBNMiner) che acquisisce conoscenza degli eventi da Internet, per la predizione dei tassi di interesse. Questo studio mostra una chiara idea di conoscenza dell'evento usando un sistema automatizzato. KBNMiner è progettato per adottare una base di conoscenza preliminare, che viene vista come "conoscenza esperta", come base su cui raccogliere informazioni sugli eventi da Internet automaticamente e quindi applicare le informazioni a un modello di Neural Network per la previsione dei tassi di interesse. Ciò che si distingue in questo studio è che la tecnica di web mining che hanno applicato per la previsione dei tassi di interesse può anche essere utilizzata per previsione del mercato azionario.

Lavori futuri

In questa sezione si sono esaminati i recenti sviluppi dei modelli di previsione del mercato azionario. Confrontando vari modelli di predizione, si è scoperto che le Neural Network offrono l'abilità prevedere le direzioni del mercato in modo più accurato rispetto ad altre tecniche esistenti. La capacità di Neural Network di imparare relazioni non lineari dalle coppie di input / output di training consente loro di modellare sistemi dinamici non lineari, come i mercati azionari, più precisamente [24].

Anche altri modelli come Support Vector Machine e Case Based Reasoning sono diventati popolari nella previsione del mercato azionario. Support Vector Machine ha mostrato la sua corretta applicazione nei problemi di classificazione e di regressione, in particolare nella previsione di serie storiche e applicazioni finanziarie correlate [32]. Case Based Reasoning è una tecnica di ragionamento che riutilizza casi passati per trovare una soluzione al nuovo problema. Case Based Reasoning cattura la conoscenza e la competenza di un'organizzazione fornendo spiegazioni per le soluzioni derivate. Per questa ragione, Case Based Reasoning è comunemente applicato a molti problemi.

Inoltre, studiando diversi punti chiave dei mercati azionari, si è scoperto che molti

²⁷Text mining è la scoperta da parte del computer di informazioni nuove, precedentemente sconosciute, estraendo automaticamente informazioni da diverse fonti di testo. È il processo di estrazione di informazioni di alta qualità dal testo. Le informazioni sono in genere derivate dall'elaborazione di modelli e pattern attraverso mezzi come l'apprendimento automatico e statistico di pattern. L'estrazione del testo di solito comporta il processo di strutturazione del testo di input (*parsing*), estrazione di pattern all'interno dei dati strutturati e infine la valutazione e l'interpretazione dell'output.

ricercatori hanno riconosciuto che fattori qualitativi come effetti politici ed eventi internazionali possono avere un impatto significativo sui prezzi delle azioni. È stato esaminato che Neural Network basate sia sui fattori quantitativi che qualitativi sono di gran lunga superiori a quelle basate solo su fattori quantitativi. Inoltre, il web è considerato come la principale fonte di notizie sugli eventi per la predizione del mercato azionario, contenendo informazioni sugli eventi più recenti e latenti. Pertanto, per prevedere il mercato, tecniche di web mining sono necessarie per avere una maggiore accuratezza e per fare predizioni in un tempo più breve.

Per ulteriori ricerche, in primo luogo, un database contenente conoscenza preliminare dovrebbe essere costruito analizzando eventi storici nei mercati azionari. Sulla base delle conoscenze precedenti, verrà sviluppato lo schema di attribuzione di peso degli eventi e ciascun evento dovrebbe essere pesato di conseguenza. Infine, il modello proposto che incorpora i pesi degli eventi nei dati numerici delle serie temporali dovrebbe essere confrontato empiricamente con altri modelli.

2.2 Sentyment

A differenza degli algoritmi appena citati, Sentyment si concentra sulla creazione di strategie di investimento, invece che la previsione dell'andamento nel tempo di un particolare titolo azionario. Gestisce contemporaneamente più asset. Non si può paragonare agli algoritmi nello stato dell'arte sul piano della precisione di previsione, perchè la funzione che cerca di massimizzare non è la stessa delle AI accademiche. Questo software commerciale ha come scopo il rispetto delle politiche scelte dall'utente in merito a propensione al rischio e interessi personalizzati e la massimizzazione del guadagno attraverso l'ottimizzazione di alcuni indicatori come, appunto, il guadagno prodotto dalla strategia o il numero di operazioni vincenti.

A differenza degli algoritmi sopra citati, in cui lo scopo è *prevedere quale valore avrà la funzione* in un certo istante, ciò che trova Sentyment è invece *quale operazione fra buy, sell o hold conviene eseguire* al fine di massimizzare il guadagno, in un certo istante.

Ciò premesso, è comunque ovvio il fatto che il sistema abbia una certa capacità di previsione del mercato, di identificazione dei pattern ricorrenti o addirittura un'integrazione di informazioni sugli eventi. Per poter competere con gli altri sistemi presenti sul mercato deve fare sicuramente uso di una o più fra le tecniche descritte in questo capitolo. Tuttavia, la AI verrà trattata come una *black-box* e non saranno quindi rivelati dettagli sulla sua implementazione, ma verranno considerati soltanto gli input e gli output.

L'analisi comparativa con altri sistemi risulta quindi difficile in quanto gli obiettivi dei software sono differenti. È comunque possibile misurare rispetto ad indicatori assoluti o comparare fra loro le diverse AI che compongono Sentyment.

2.2.1 Cenni all'architettura

Sentyment è il prodotto di un'azienda milanese che opera nel settore informatico-finanziario. Nexid Edge è un ramo del gruppo NEXiD votato per perseguire l'innovazione tecnologica e sviluppare soluzioni di digital business rivoluzionarie. Sentyment è la nuova Business Unit dedicata all'intelligenza artificiale.

L'obiettivo è offrire alle organizzazioni un accesso senza pari a tecnologie all'avanguardia che uniscono il meglio della tecnologia e dell'imprenditorialità per elevare la saggezza collettiva. Al momento, grazie all'intensa collaborazione con Nexid Finance, Nexid Edge è attiva nel fornire progetti di consulenza digitale end-to-end, per offrire ai clienti un'esperienza di *augmented consulting*.²⁸

²⁸Nexid, Srl. URL: <https://nexid.it/>

In questo documento, tuttavia, si fa riferimento a "Sentymment" intendendo l'applicazione sviluppata e non la divisione AI di Nexid.

Essendo il fine ultimo della tesi l'inserimento un nuovo modulo operativo in Sentymment, è necessario prendere familiarità con il software e i dati che tratta.

L'applicazione di intelligenza artificiale era già esistente e in grado di raccogliere dati da Kraken e produrre le strategie ma mancava di un'architettura modulare, di componenti separati che assolvono a ruoli precisi e definiti e di un'interfaccia ad API per interrogare il sistema in esecuzione. Era quindi necessario sviluppare da zero *data module* e *coordinator* e permettere l'integrazione del resto del sistema già esistente nella nuova architettura a moduli. Era inoltre richiesto lo sviluppo della nuova AI di controllo, supervisore che sceglie la migliore fra le AI in esecuzione.

Tutti i moduli del sistema sono scritti in Python3, comprese le AI stesse. Alcune loro parti sono state riscritte in Cython e Go per migliorare le prestazioni.

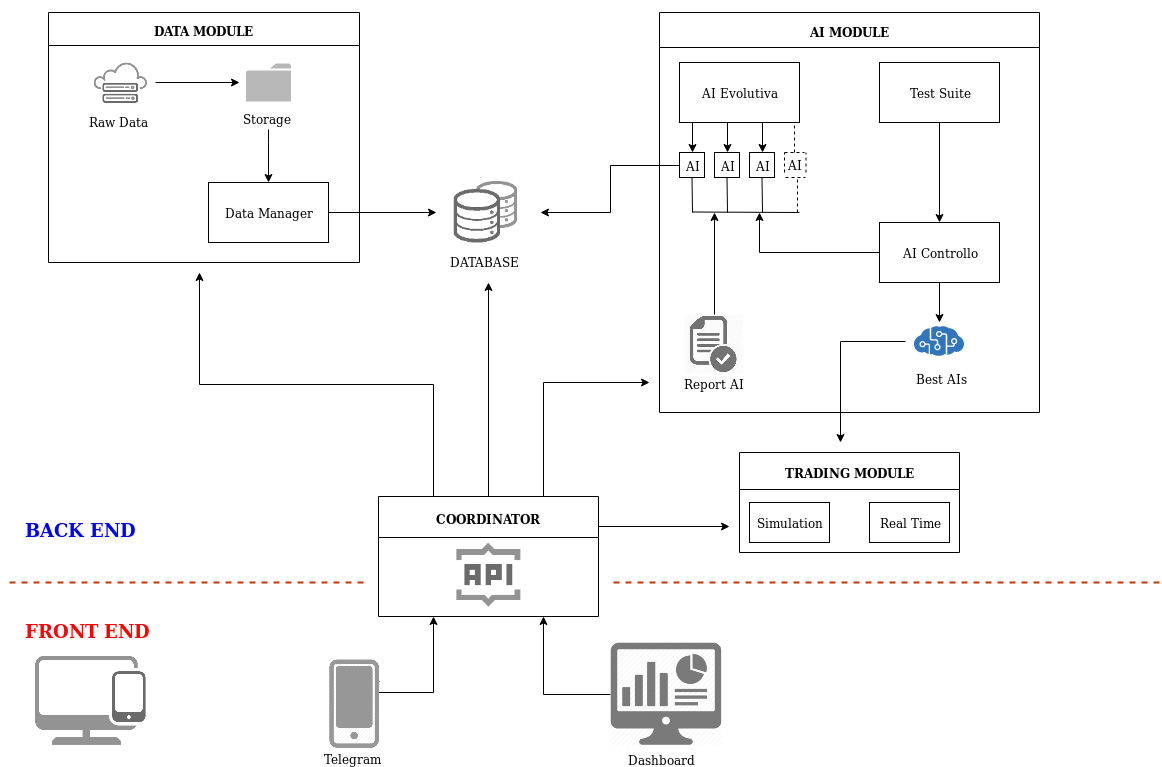


Figura: Architettura di Sentymment. Nei riquadri sono mostrati i quattro principali componenti, suddivisi al loro interno in ulteriori sotto-moduli. Il grafico è diviso in *front-end* e *back-end*, che rappresentano rispettivamente le interfacce esposte all'utente utilizzatore del sistema, e la parte di logica non accessibile dall'esterno

Verranno ora brevemente presentati i componenti di Sentyment.

- *data module* raccoglie i record di trade raw tramite le API Kraken, gestisce la creazione e scrittura delle candele OHLCV e permettere l'accesso a tutti i dati e i database da parte degli altri componenti. I sotto-moduli si spartiscono i compiti: quando i dati di trade sono scaricati e salvati su uno storage temporaneo da *raw data*, allora *data manager* li copia sul database permanente e in automatico scatta la creazione delle nuove candele OHLCV, che sono create allo scoccare della nuova ora a partire dai dati appena scaricati (o minuto/giorno, a seconda delle configurazioni scelte). Infine le nuove candele create sono inserite nel database.
- *coordinator* implementa le interfacce esposte per il controllo e l'interrogazione dello stato di Sentyment. È l'unico punto di accesso al sistema in esecuzione. Tramite delle API esposte su server è possibile conoscere quali asset sono attualmente in fase di scaricamento e le informazioni riguardo le configurazioni delle candele che si vogliono creare; è inoltre possibile impostare i timeframe delle candele (orarie, giornaliere), attivare o disattivare i download dati e forzare la creazione di specifiche candele.
- *AI module* è la parte preesistente di intelligenza artificiale. Contiene le numerose AI che creano le strategie di investimento e il supervisore che sceglie la migliore fra queste, a intervalli fissati. È inoltre presente un modulo che crea report che descrivono l'andamento delle AI, i dati che stanno processando e le statistiche calcolate per ognuna di esse, per misurarne la performance online.
- *trading module* si occupa di seguire le strategie create dalle AI per effettuare direttamente operazioni di acquisto e vendita tramite le API di Kraken. È collegato ad un portafoglio Kraken che contiene un certo budget ed è libero di fare trading. Oltre ad agire sul mercato vero, le AI operano anche in un ambiente simulato che permette di riprodurre e monitorare quale sarebbe stato il loro comportamento nel mercato reale o in nuovi mercati non ancora raggiunti, al fine di calcolare ulteriori statistiche e tracciare l'evoluzione delle AI.

Come già accennato, sono stati sviluppati *data module* e *coordinator*, ridisegnando la preesistente architettura di Sentyment e integrando gli altri componenti. Durante la riprogettazione si sono potute scegliere liberamente la logica di implementazione e

le tecnologie da impiegare, con la restrizione di dover utilizzare Python3, con *sqlalchemy*²⁹ come *ORM*³⁰ e *flask*³¹ per implementare il server e le API.

Prima di sviluppare la AI supervisore sono state testate le AI singolarmente attraverso alcune metodologie di test che saranno descritte di seguito. È stato svolto un lavoro di analisi dei dati prodotti dalle AI, confrontandoli con alcune strategie di base e con altre intelligenze artificiali sviluppate al fine di misurarne le prestazioni e solo dopo le AI sono state confrontate fra di loro. I metodi ritenuti migliori per il confronto sono stati scelti in seguito ad un'analisi del contesto e alla natura finanziaria dei dati.

2.2.2 Strumenti e tecnologie utilizzati

Sentymment è quasi totalmente scritta in Python3, tranne per alcune parti in cui si ha bisogno di elevate performance che sono quindi state riscritte in Cython³². Queste comprendono, ad esempio, gran parte del codice delle AI, ma non solo.

Si fa uso estensivo di librerie di analisi dati quali *numpy*³³, *pandas*³⁴ e *matplotlib*³⁵ per visualizzare i grafici dei dati, soprattutto nella parte di sviluppo e analisi.

Queste librerie sono i classici strumenti utilizzati nell'analisi dati con Python e sono i più utilizzati; le alternative attualmente presenti non offrono le stesse funzionalità con la medesima efficienza e per questo erano anche i *tool* già usati nella prima versione di Sentymment.

Pandas, in particolare, permette di importare grandi quantità di dati direttamente

²⁹SQLAlchemy è un toolkit SQL open-source e object-relational mapper (ORM) per Python

³⁰L'Object-Relational Mapping (ORM) è una tecnica di programmazione che favorisce l'integrazione di sistemi software aderenti al paradigma della programmazione orientata agli oggetti con sistemi RDBMS (Relational Database Management System). Un prodotto ORM fornisce, mediante un'interfaccia orientata agli oggetti, tutti i servizi inerenti alla persistenza dei dati, astruendo nel contempo le caratteristiche implementative dello specifico RDBMS utilizzato.

³¹Flask è un micro framework web scritto in Python. Ha un nucleo semplice ma estendibile: ci sono ad esempio estensioni per la validazione delle form, la gestione del caricamento dei file, varie tecnologie di autenticazione ed altro

³²Cython è un linguaggio di programmazione che mira ad essere un superset del linguaggio di programmazione Python, progettato per fornire prestazioni di tipo C con codice scritto principalmente in Python con sintassi aggiuntiva opzionale ispirata al C

³³NumPy è una libreria open source per il linguaggio di programmazione Python, che aggiunge supporto a grandi matrici e array multidimensionali insieme a una vasta collezione di funzioni matematiche di alto livello per poter operare efficientemente su queste strutture dati.

³⁴Pandas è una libreria software scritta per il linguaggio di programmazione Python per la manipolazione e l'analisi dei dati. In particolare, offre strutture dati e operazioni per manipolare tabelle numeriche e serie temporali. Il nome deriva dal termine "panel data", termine econometrico per set di dati che include osservazioni su più periodi di tempo per gli stessi individui.

³⁵Matplotlib è una libreria per la creazione di grafici per il linguaggio di programmazione Python e la libreria matematica NumPy. Fornisce API orientate agli oggetti che permettono di inserire grafici all'interno di applicativi

da database o da file ed è in grado di gestirli in modo efficiente indicizzandoli e garantendo un accesso rapido. Un ampio set di operazioni statistiche sui dati lo rende uno strumento indispensabile in *data science* e ampiamente utilizzato. Basti pensare ad un AI a cui servono in input moltissimi dati: quando questi sono salvati su database è facile eseguire piccole letture e query puntuali e aggregate ma quando, invece, si ha bisogno di portarne in memoria un gran numero per processarli con funzioni al di là delle potenzialità di *SQL*, serve un ulteriore strumento capace di gestire l'accesso efficiente alla memoria e che offra direttamente queste funzioni di analisi (esempio con funzione *resample* nel Capitolo 3).

Il database scelto è di tipo *relazionale*. Si è scelta l'opzione relazionale invece di *noSQL* in quanto i dati sono tutti perfettamente strutturati (schema dati in dettaglio nel Capitolo 3): la sorgente principale da cui partono tutte le elaborazioni è rappresentata dai dati di trading *raw* forniti da Kraken. Questi hanno ciascuno un riferimento al *timestamp*, alla tipologia dell'operazione di trade, la quantità di titoli venduti / acquistati e il loro volume. Sostanzialmente non servono altre informazioni per la creazione delle candele OHLCV, che sono anch'esse strutturate ed esplicitano *open*, *high*, *low*, *close*, *volume*, *timestamp* finale e altri dati secondari. Le AI lavorano tutte sulle candele e necessitano soltanto degli attributi appena citati.

L'unica eccezione, in cui dati non sono perfettamente strutturati, sono le AI. Sono memorizzate come un insieme di parametri che permettono la loro ricreazione in un secondo momento, se necessaria, e sono salvati in formato *JSON*.³⁶ Dato che i parametri sono diversi, non strutturati e con possibilità di molteplici valori nulli, la memorizzazione in JSON già adottata nella prima versione di Sentyment sembrava comunque adatta, permettendo una compressione più efficiente dei dati e un accesso orientato al documento. Di una specifica AI si vogliono conoscere i suoi parametri, ma il nome / *id* della AI è noto: per esempio, in seguito ad un'analisi durante un nuovo ciclo di test / sviluppo, si vuole continuare dalla AI più promettente fra quelle della precedente iterazione e quindi si accede al documento che contiene tutti i parametri della AI scelta.

I JSON dei parametri delle AI non sono scritti su database, ma sono dei file che risiedono direttamente su file system nelle directory di Sentyment, strutturate in modo da poter accedere ai parametri di una certa AI usando il suo identificativo. In sviluppi futuri potrebbe essere più adatto l'uso di un database non relazionale per la memorizzazione delle AI, oppure sfruttare i tipi di dato JSON presenti in ormai tutti i RDBMS moderni e quindi salvare tutto nello stesso database.

³⁶JSON, acronimo di JavaScript Object Notation, è un formato adatto all'interscambio di dati fra applicazioni client/server. È basato sul linguaggio JavaScript ma ne è indipendente.

Fra i database relazionali è stato scelto *Mysql*³⁷ perché meglio si sposa con l'ORM di *sqlalchemy*.

Sqlalchemy permette di manipolare database SQL direttamente da Python. Mappa lo schema tabellare dei dati in uno a oggetti più facilmente processabile dal linguaggio di programmazione scelto, in questo caso Python. Grazie alla tecnologia degli ORM, il modello dei dati è sempre disponibile come classe del linguaggio e risulta quindi più comodo il loro accesso e manipolazione: ora si lavora con collezioni di oggetti invece che con un semplice elenco di record contenenti alcuni campi. L'ORM mette anche a disposizione un motore di astrazione per le *query*. Tramite un modello ad API le query eseguite in Python sono tradotte nel dialetto SQL del DBMS sottostante. Al programmatore resta il compito di scrivere le query usando l'*expression language*³⁸ di sqlalchemy, il sistema usato per rappresentare strutture dei database relazionali e espressioni usando Python.

Flask, infine, è un microframework per Python ed implementa *coordinator*, il web server che espone le informazioni disponibili sullo stato di esecuzione di Sentyment. È molto minimalista se mantenute le impostazioni di default e serve soltanto documenti JSON e non pagine HTML.

In futuri sviluppi le informazioni fornite dal web server in JSON potrebbero essere processate da un ulteriore componente, che le esporrebbe in modo più user-friendly attraverso un software di front-end e permetterebbe l'interrogazione di *coordinator*, attraverso le API, in maniera più intuitiva.

³⁷MySQL o Oracle MySQL è un Relational database management system (RDBMS) composto da un client a riga di comando e un server. È un software libero rilasciato a doppia licenza ed è sviluppato per essere il più possibile conforme agli standard ANSI SQL e ODBC SQL. I sistemi e i linguaggi di programmazione che supportano MySQL sono molto numerosi.

³⁸Expression Language (EL) è un linguaggio per creare una rappresentazione interpretabile di conoscenze specifiche. In questo caso è un linguaggio creato da sqlalchemy appositamente per rappresentare query SQL nel linguaggio di Python.

Capitolo 3

Elaborazione dati di trading

In questo capitolo si andrà principalmente a descrivere il lavoro svolto dal software di raccolta e gestione dati, *data module*, e alcuni dettagli della sua implementazione. Attraverso l'analisi delle operazioni svolte dal programma si darà anche un ulteriore sguardo ai dati, a come sono organizzati dalla piattaforma di trading e all'interno del software di Nexid, e a come si processano per ottenere un formato utile per l'elaborazione da parte delle AI.

Anche se Sentyment comprendeva già una parte di logica di scaricamento dati, in seguito alla sua riorganizzazione è stato necessario riscrivere questa parte di software convertendola interamente in un modulo separato, sempre attivo e pronto a scaricare, con cui si può interagire attraverso *coordinator*. Svolge diverse attività fra cui scaricare grazie alle API di Kraken, creare le candele OHLCV a partire dai dati scaricati e gestire il database.

I requisiti dell'applicativo sono i seguenti:

- Scaricare dati di trade *raw* per diversi tipi di asset, restando aggiornati in tempo reale con l'arrivo di nuovi
- Fornire la possibilità di scaricare dati passati
- Creare candele OHLCV a partire dai dati scaricati e poter sceglierne i parametri (lunghezza candela, periodo da considerare)

3.1 Download dati Kraken

L'input di tutto il sistema sono i dati di trade. Un modulo intero è dedicato alla logica di scaricamento dati, il cui fine è soltanto fornire i suddetti dati alle intelligenze artificiali, che rappresentano la parte principale del software. Tutta l'architettura

è orientata al dato; le AI stesse ignorano l'esistenza di *data module*, perché di fatto accedono direttamente al database per prelevare i dati, ignorando chi lo abbia riempito. Un metodo strutturato per "riempire" il database e preparare i dati per le AI è quindi un passo iniziale e fondamentale per tutta l'architettura.

Le API di Kraken si dividono in *pubbliche* e *private*: tutte le attività svolte da *data module* fanno uso soltanto delle interfacce pubbliche, mentre quelle private, le quali necessitano di un API Key¹, servono per collegare un portafoglio con budget e permettono di eseguire operazioni di acquisto e vendita di titoli. Sono utilizzate soprattutto dal modulo di trading.

Le API pubbliche permettono molti tipi di operazioni fra cui, le più usate:

- **Get tradable asset pairs:** elenco di tutti gli asset trattati da Kraken, i *ticker*², con associato il nome usato e altre informazioni. Sono 110 in totale le coppie di valute di scambio fra una criptovaluta e la sua controparte in Euro o Dollaro. Alcuni esempi:
 - XXBTZEUR: ticker che rappresenta la valuta Bitcoin-Euro (XBT-EUR)
 - XXBTZUSD: Bitcoin-Dollaro (XBT-USD)
 - XETHZEUR: Ethereum-Euro (ETH-EUR)
 - XETHZUSD: Ethereum-Dollaro (ETH-USD)

Non ci sono soltanto valute di scambio fra Euro / Dollaro e una criptovaluta, ma bensì anche scambi fra criptovaluta e criptovaluta, come, ad esempio:

- ADAETH: valuta di scambio fra le due criptovalute Cardano (ADA)³ ed Ethereum (ETH)

¹Una chiave per API (API Key) è un identificatore univoco utilizzato per autenticare un utente, uno sviluppatore o il programma chiamante un'API. Tuttavia, vengono in genere utilizzati per autenticare un progetto con l'API anziché un utente umano. Spesso si usano per accedere a quelle API che forniscono servizi a pagamento o per superare limitazioni di utilizzo.

²Il ticker è un'abbreviazione che identifica le società che vengono quotate su un mercato finanziario. Quando una società decide di 'diventare pubblica', sceglie il tipo di mercato sui cui essere quotata e il ticker di identificazione. Il ticker viene chiamato anche ticker azionario oppure ticker simbolo. Il termine deriva in effetti dal 'ticker tape' o 'nastro per telescrivente', un nastro su cui, nel passato, si trasmettevano le informazioni relative ai prezzi.

³Cardano è una piattaforma che gestisce la criptovaluta ADA. Il progetto è guidato dall'ex cofondatore di Ethereum. Lo scopo di Cardano è creare una piattaforma adatta allo sviluppo di smart contracts e un sistema di pagamento (altro sulle criptovalute nel Capitolo 3).

– XZECXXBT: Zcash (ZEC)⁴ e Bitcoin (XBT)

Gli asset che scambiano due criptovalute fanno parte dei ticker gestiti da Kraken e quindi sono comunque scaricati e memorizzati nel database, ma non verranno usati nel resto del progetto.

L'elenco è scaricato in una fase iniziale di setup del modulo dati e ha lo scopo di memorizzare su database i nomi usati da Kraken per tutti gli asset che fornisce, in modo da poterli utilizzare nelle altre API per la fase successiva, in cui si richiede l'elenco di trade di uno specifico ticker.

Come si può notare sono usati nomi particolari per rappresentare i ticker ed è quindi necessario memorizzare il mapping fra nome reale e nome usato da Kraken per poi poter utilizzare le API che seguono.

- **Get recent trades:** Ora che si conoscono i nomi dei ticker usati da Kraken si possono usare per richiedere, per ciascuno di essi, l'elenco delle transazioni di mercato. Questa è in assoluto la API più usata dal modulo dati: riceve i dati di trade *raw* che costituiscono l'input principale da cui partono le successive elaborazioni. I dati di trade sono memorizzati e saranno usati per creare le candele OHLCV, l'input per le intelligenze artificiali. Attraverso questa API è possibile sia richiedere dei dati storici in un certo periodo di tempo passato, sia ricevere aggiornamenti sui nuovi dati disponibili. In entrambi i casi va specificato come parametro il ticker di cui si vuole conoscere l'elenco degli scambi. Come secondo parametro è necessario un riferimento temporale che indichi quali trade sono richiesti di preciso (parametro *since*).

Esempio di documento JSON restituito dalla chiamata a questa interfaccia, usando come parametri il ticker XXBTZEUR e il riferimento a gennaio 2020:

```
{
  "error": [],
  "result":
  {
    "XXBTZEUR":
    [
      ["6410.20000", "0.07588465", 1577833204.9008, "b", "m", ""],
```

⁴Zcash è una criptovaluta che offre privacy e trasparenza selettiva delle transazioni. I pagamenti Zcash sono pubblicati su una blockchain pubblica, ma il mittente, il ricevente e il valore della transazione possono rimanere privati. Come Bitcoin Zcash ha una fornitura fissa totale di 21 milioni di unità.


```

["6410.00000","0.02911527",1577833220.5154,"s","1",""],
["6410.00000","0.02910911",1577833261.2323,"s","1",""],
["6435.50000","0.08000000",1577842069.8177,"s","1",""], ...
]
"last":"1577842069817694680"
}
}

```

Oltre ad alcuni metadati utili per riutilizzare le API, sono principalmente elencati una serie molto lunga di record formati da sei campi: *price*, *volume*, *time*, *buy/sell*, *market/limit*, *miscellaneous*.

Prendendo come esempio il primo record:

- *price*: 6410.20000. Siccome la valuta è Bitcoin-Euro, il numero indica il valore di un titolo Bitcoin espresso in Euro.
- *volume*: 0.07588465. La quantità di titoli Bitcoin scambiati in quell'istante.
- *time*: 1577833204.9008. Timestamp associato all'operazione in formato unix timestamp.⁵
- *buy/sell*: "b". Indicazione del tipo di operazione: 'b' (buy) o 's' (sell).
- *market/limit*: "m". Indica se comprare a prezzo di mercato (m) o se applicare un limit. Un limit order è un ordine inserito nell'order book con un prezzo limite specifico. Quando è inserito un limit order, l'operazione verrà eseguita solo se il prezzo di mercato raggiunge il prezzo limite. È quindi possibile usare limit order per comprare a un prezzo più basso o per vendere a un prezzo più alto rispetto all'attuale prezzo di mercato.
- *miscellaneous*: "". Spazio per eventuali commenti o casi particolari.

Il campo *last* serve da passare come nuovo parametro *since* nella prossima chiamata, a indicare di scaricare ancora i dati dal punto in cui ci si era fermati. Siccome il volume dei dati di trade è molto elevato,⁶ viene diviso in frammenti; per continuare a scaricare frammento per frammento si sfrutta la combinazione di questi due parametri.

- **Get OHLC data:** Kraken offre anche la possibilità di scaricare direttamente le candele OHLCV create a partire dai dati di trade raw, ma questa funzionalità è già implementata all'interno di Sentymment stesso. Ad ogni modo, sono

⁵Nei sistemi operativi Unix e Unix-like il tempo viene rappresentato come offset in secondi rispetto alla mezzanotte del 1° gennaio 1970. I decimali rappresentano la frazione di secondo.

⁶I dati di trade di Bitcoin per il solo anno 2019 hanno un'occupazione dell'ordine di grandezza dei 10 Gb.

comunque riportati alcuni esempi di dati scaricati tramite questa API.

Esempio di documento JSON restituito dalla chiamata, con parametri il ticker XXBTZEUR, il timestamp di gennaio 2020 e la lunghezza di candela a un minuto:

```
{
  "error": [],
  "result":
  {
    "XXBTZEUR":
    [
      [1582093500, "9384.7", "9384.7", "9384.6", "9384.6", "9384.6",
        "0.35398923", 9],
      [1582093560, "9384.6", "9384.7", "9384.0", "9384.0", "9384.5",
        "0.04147968", 5],
      [1582093620, "9384.1", "9384.1", "9384.0", "9384.0", "9384.0",
        "0.40413908", 5],
      [1582136640, "9433.6", "9433.6", "9433.5", "9433.5", "9433.5",
        "0.33265963", 4], ...
    ],
    "last": 1582136580
  }
}
```

La struttura del documento è simile a quella restituita nell'esempio precedente, con la differenza che ora i record sono composti da attributi diversi. Prendendo come esempio il primo record:

- *time*: 1582093500. Timestamp finale della candela. La candela delle 12:00 comprende i trade dalle 11:00 alle 11:59.
- *open, high, low, close*: rispettivamente 9384.7, 9384.7, 9384.6, 9384.6. Prezzo di apertura e chiusura della candela, minimo e massimo.
- *vwap*: 9384.6. Prezzo medio della candela.⁷
- *volume*: 0.35398923. Somma totale della quantità di titoli scambiati durante l'intervallo di tempo della candela.
- *count*: 9. Numero di operazioni buy / sell realmente eseguite all'interno della durata della candela.

⁷In finanza, Volume Weighted Average Price (VWAP) è il rapporto tra il valore scambiato e il volume totale scambiato in un determinato orizzonte temporale. È una misura del prezzo medio al quale un titolo viene negoziato nell'orizzonte di negoziazione.

Il significato e la definizione degli attributi delle candele verranno trattati in dettaglio più avanti.

Il modulo di scaricamento è molto semplice e interroga solamente le API di Kraken, memorizzando di volta in volta blocchi di dati su database. L'unica difficoltà incontrata risulta essere la comprensione di alcuni parametri delle API, che si dimostrano non perfettamente documentate e pertanto è lasciato ai programmatori scoprire il significato di alcune interfacce.

Per esempio il parametro *last*, che dovrebbe indicare l'ultimo timestamp fra i record ritornati nella chiamata e che va passato come parametro *since* nella successiva chiamata, per ripartire dall'ultimo trade, è controverso. La documentazione riporta "*return trade data since given id*", intendendo quindi che si tratti di un identificatore: curiosamente in moltissimi casi (soprattutto nei dati OHLC, che hanno tutti timestamp preciso all'ora) ha lo stesso valore del timestamp e quindi questo significa che i timestamp sono usati come id. Quando, però, il valore ha dei decimali, che nel caso di un unix timestamp indicano frazioni di secondo, questi sono aggiunti in coda all'identificatore con un *padding*⁸ a 0 di nove cifre. Quando si richiedono i dati a partire da un certo timestamp che non ha decimali, quindi, bisogna aggiungere il padding altrimenti l'id richiesto non viene trovato nei database di Kraken. In risposta è invece ritornano il record con id minore di tutti, che è quello che più si avvicina al valore richiesto, estremamente basso rispetto a tutti gli altri id presenti nel database dato che ha nove cifre mancanti.

Tutte queste informazioni si possono inferire analizzando molte chiamate con argomenti particolari ma non vi è una documentazione completa che espone chiaramente l'uso corretto del parametro e il risultato sono dei comportamenti inaspettati nei programmi che usano le API. Una volta capito il problema è facile aggirarlo, ma la scelta progettuale di utilizzare i timestamp convertiti a id è discutibile e inoltre non chiaramente documentata.

3.2 Creazione candele OHLCV

Una volta memorizzati tutti i dati di trade raw e disponibili per le letture, è il momento di creare il dato aggregato che riassume le operazioni di scambio e che rappresenta la principale fonte di input per le intelligenze artificiali.

⁸Il padding è una delle numerose pratiche che prevedono l'aggiunta di dati all'inizio, a metà o alla fine di un messaggio. È usato per rendere uniforme un certo dato, allineandolo alla lunghezza specificata: se la lunghezza del messaggio finale deve assumere precisamente un certo valore ma il testo contenuto è più breve, si aggiungono tanti valori di padding quanti la differenza fra la lunghezza del messaggio da inviare e quella del testo contenuto.

Le candele sono interamente generate grazie alla libreria di Python *pandas*, chiamando il metodo *resample*⁹:

```
def resample(self, rule, ... ):
    Resample time-series data.
```

```
Convenience method for frequency conversion and resampling of time series.
Object must have a datetime-like index (DatetimeIndex, PeriodIndex,
or TimedeltaIndex), or pass datetime-like values to the on or level keyword.
```

È un metodo della classe *DataFrame*, quindi viene chiamato a partire da un oggetto contenente il dataset da ricampionare per creare la candela; come parametro è usato *rule*, un *timedelta* che rappresenta la lunghezza della candela.

Supponendo di avere dunque tutti i dati raw disponibili nel database e di voler creare candele con timeframe di un'ora, per tutto il mese di gennaio 2020, il procedimento seguito è il seguente:

Dal punto di vista concettuale, il metodo citato aggrega i dati creando un nuovo record contenente i quattro attributi *open*, *high*, *low*, *close*, *volume*. Formalmente:

- *open*
- *high*
- *low*
- *close*
- *volume*: sommatoria singoli volumi

Dove, per ogni record, *price* è il prezzo dell'asset scambiato, *volume* è....

3.3 Indicatori

3.4 Creazione strategie

⁹Da [documentazione pandas:](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.resample.html)

[https://pandas.pydata.org/pandas-](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.resample.html)

Capitolo 4

Test per intelligenze artificiali

I test sono in realtà un elemento chiave per il funzionamento dei progetti di intelligenza artificiale. Non si sviluppa semplicemente un algoritmo AI fornendo dati di training e mettendolo in produzione; è necessario verificare effettivamente che i dati di addestramento svolgano un lavoro sufficientemente buono per classificare accuratamente con una generalizzazione sufficiente senza incorrere in *overfitting* o *underfitting*. Questo viene realizzato usando tecniche di validazione e mettendo da parte un sottoinsieme dei dati di training da usare durante la fase di validazione. In sostanza, si tratta di una sorta di test QA in cui si vuole assicurare che l'algoritmo e i dati, oltre a iperparametri ¹ e metadati associati, lavorino tutti insieme per fornire i risultati predittivi sperati.

Se si sbaglia nella fase di validazione, bisognerebbe tornare indietro, cambiare i parametri e ricostruire di nuovo il modello, magari con dati di training migliori. Fatto ciò, si torna indietro e si utilizzano nuovi casi di test per verificare che il modello funzioni davvero come dovrebbe. Sebbene sembrino tutti aspetti del test e della validazione, questo accade durante la fase di addestramento della AI, prima che il modello sia messo in funzione.

Anche in fase di training, si stanno testando diversi aspetti. Innanzitutto, bisogna assicurarsi che l'algoritmo AI stesso funzioni. Non ha senso modificare gli iperparametri e allenare il modello se l'algoritmo è implementato in modo errato. Tuttavia, in realtà, è difficile avere un algoritmo non corretto perché la maggior parte di questi sono già inseriti nelle varie librerie AI: se si necessita di *K-Means Clustering*, *Support Vector Machine* o diversi tipi di reti neurali, basta semplicemente chiamare quella

¹Iperparametri sono quei parametri i cui valori sono decisi prima che inizi il processo di training. Si differenziano dai parametri che sono invece ricavati dal training stesso e vanno a comporre i settaggi dell'algoritmo AI. Gli iperparametri sono invece parametri del modello di training

funzione di libreria in Python *scikit-learn* o qualunque sia lo strumento scelto. Gli sviluppatori AI non dovrebbero scrivere gli algoritmi da zero a meno non si abbia davvero una buona ragione per farlo; ciò significa che se non li si sta codificando da zero, non c'è molto da testare per quanto riguarda la correttezza del codice reale - si suppone che gli algoritmi abbiano già superato i loro test, per quanto possibile. In un progetto AI, il QA non sarà mai focalizzato sull'algoritmo AI stesso o sul codice, supponendo che sia stato implementato come dovrebbe.

Restano dunque due cose da testare nella fase di addestramento per il modello AI stesso: i dati di training e i le configurazioni degli iperparametri. In quest'ultimo caso, è possibile testare mediante l'uso di metodi di validazione, tra cui *k-fold cross validation* e altri approcci. Ciò contribuirà a determinare se le impostazioni dell'iperparametro sono corrette.

Pertanto, tutto quello che rimane da testare sono i dati stessi per il modello AI. Ciò significa non solo qualità dei dati, ma anche completezza. Il modello di formazione rappresenta adeguatamente la realtà di ciò si sta cercando di generalizzare? Si è inavvertitamente incluso del *bias* informativo o indotto dall'uomo nei dati di training? Si sta ignorando parte dei dati che funzionano in allenamento ma falliranno durante la predizione perchè i dati del mondo reale sono più complessi? Il QA per il modello di intelligenza artificiale ha a che fare con la garanzia che i dati di training includano un campione rappresentativo del mondo reale ed eliminino il più possibile *bias* umani.

Un sistema ben validato e ben generalizzato che utilizza dati di addestramento rappresentativi e algoritmi da una fonte già testata e comprovata dunque dovrebbe dare i risultati previsti. Ma cosa succede quando non si ottengono quei risultati? La realtà è ovviamente più complessa. Nel mondo reale accadono cose che non avvengono nell'ambiente di test e può accadere che nella fase di "inferenza", quando il modello è reso operativo, non si incontrino i risultati sperati.

I problemi che sorgono con i modelli nella fase di inferenza sono quasi sempre problemi di dati o disallineamenti nel modo in cui il modello è stato addestrato rispetto ai dati del mondo reale. Se si è certi che l'algoritmo funziona e che i dati del modello di training e gli iperparametri sono stati configurati al meglio, significa che quando i modelli falliscono, si hanno problemi di disadattamento dei dati o del mondo reale. Se dei dati di input sono errati, è necessario trovarli ed analizzarli. Se il modello non sta generalizzando bene, se c'è qualche sfumatura dei dati che deve essere aggiunta per addestrare ulteriormente il modello, allora è necessario passare attraverso un ciclo completamente nuovo di sviluppo di un modello di intelligenza artificiale con nuovi dati di addestramento e configurazioni di iperparametri per affrontare il giusto livello di adattamento a tali dati. Indipendentemente dal problema, le organizzazioni che rendono operativi i modelli di intelligenza artificiale hanno bisogno di un approccio

solido in base al quale possono tenere sotto controllo le prestazioni dei modelli e controllare la versione di quelli in funzione.

I progetti di intelligenza artificiale sono davvero unici in quanto ruotano attorno ai dati. I dati sono l'unica cosa che nei test è garantita crescere e cambiare continuamente. Pertanto, è necessario considerare le AI come anche in continua crescita e cambiamento. [43]

4.1 Test per decision support AI

4.2 Test framework workflow

4.3 Analisi offline: maximum profit e genetic AI

4.4 Analisi online: reinforcement learning

Verrà ora affrontato il tema centrale della tesi: sviluppare una AI supervisore in grado di scegliere un rappresentante migliore fra una serie di agenti intelligenti.

Sentiment è composta da numerose AI che creano strategie di investimento e ognuna opera indipendentemente dalle altre, producendo costantemente dati relativi ai nuovi trade aggiornati in tempo reale e differenziandosi leggermente dalle altre. Le AI attive sono solitamente 12, ma il numero potrebbe variare a seconda degli iperparametri scelti durante il ciclo di test. Per ogni AI sono noti i suoi *trigger*, cioè le operazioni di buy / sell relative ad un certo istante di tempo che l'agente ha prodotto.

time	trigger
2020-01-13 21:00:00	1
2020-01-13 22:00:00	0
2020-01-13 23:00:00	0
2020-01-13 00:00:00	0
2020-01-13 00:00:00	0
2020-01-13 01:00:00	-1

Tabella: elenco di record che rappresentano il risultato prodotto da una delle AI di Sentiment. Sono state considerate le candele orarie: per ogni ora si ha il segnale di buy / sell / hold. I trigger sono rappresentati da un valore numerico: 0 per hold, 1 per buy e -1 sell.

Inizialmente si può supporre di raccogliere dati per alcune settimane e poi utilizzare questo dataset per scegliere una AI, quindi ricominciare la fase di osservazione, in cui si raccolgono dati per alcune settimane, e ripetere il ciclo; in una fase più avanzata, però, anche il tempo lungo cui prendere le misure diventa un parametro del supervisore e deve essere deciso. Unendo l'elenco di operazioni di ogni AI per un certo periodo con il dataset OHLCV dei prezzi, è possibile calcolare uno *score* basato su numerose statistiche. La AI supervisore, fondandosi su questo score, sceglie qual è la AI che lo massimizza di più rispetto alle altre, tenendo conto anche del fatto che il punteggio varia a seconda del periodo di tempo considerato e anche all'interno dello stesso dataset, se si scelgono punti iniziali e finali differenti.

Un approccio di machine learning che si adatta particolarmente al contesto descritto è il *reinforcement learning*, tecnica di apprendimento automatico che punta a realizzare agenti autonomi in grado di scegliere azioni da compiere per il conseguimento di determinati obiettivi tramite interazione con l'ambiente in cui sono immersi. L'apprendimento per rinforzo [44] è uno dei tre paradigmi principali dell'apprendimento automatico, insieme a *supervised* e *unsupervised learning*. A differenza degli altri due, questo paradigma si occupa di problemi di decisioni sequenziali, in cui l'azione da compiere dipende dallo stato attuale del sistema e ne determina quello futuro. La qualità di un'azione è data da un valore numerico di "ricompensa", ispirata al concetto di reinforcement, che ha lo scopo di incoraggiare comportamenti corretti dell'agente. Questo tipo di apprendimento è solitamente modellato tramite i processi decisionali di Markov e può essere effettuato con diverse tipologie di algoritmi.

Questa tecnica si basa sul presupposto che all'interno di un sistema si possano: scegliere degli output sulla base degli input ricevuti, valutare l'efficacia degli output rispetto ad un preciso parametro di riferimento e cambiare il meccanismo di scelta degli input per massimizzare la valutazione di efficacia.

Quando si effettua una scelta efficace allora la valutazione di efficacia manda in output un premio proporzionale all'efficacia della scelta. Quando invece è effettuata una scelta inefficace allora l'output della valutazione manda una penalità proporzionale. Osservando le scelte e le ricompense, si cerca di modificare la funzione matematica che regola le scelte in modo da massimizzare la quantità e la qualità dei "premi".

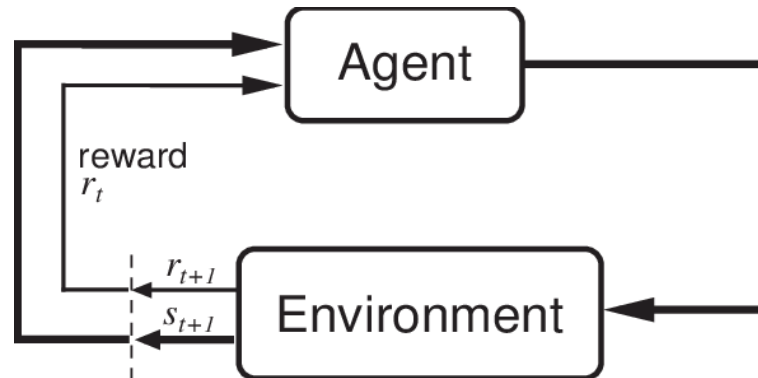


Figura: Idea generale di *reinforcement learning*: un agente interagisce con l'ambiente per imparare dall'esperienza. Per ogni istante di tempo t l'agente è in un certo stato S_t ed esegue l'azione A_t . Come risultato, passa in un nuovo stato S_{t+1} e riceve una ricompensa R_{t+1}

Nel contesto di Sentyment il supervisore rappresenta l'agente che opera le scelte. L'ambiente (i dati in input) sono i dataset di qualche settimana che, a ogni iterazione del ciclo di scelta della nuova AI rappresentante, sono forniti all'algoritmo di reinforcement learning. La scelta che il supervisore deve operare ad ogni passo è decidere quale fra le dodici AI sta, secondo lui, agendo meglio delle altre; mentre l'output della scelta, ovvero la valutazione di efficacia, è il risultato di una funzione di score calcolata con indicatori statistici finanziari a partire dai trigger della AI scelta, il cui risultato viene confrontato con quello prodotto dalle altre AI competitori.

4.4.1 Q-learning

4.4.2 Multi-armed bandit

Un problema noto che si adatta alla situazione e fa uso di tecniche di reinforcement learning, implementato usando algoritmi di q-learning. La sua versione classica discosta leggermente da questo ambito ma può essere revisionato per trarre vantaggio dalla sua semplicità.[44]

Il problema originale illustra un contesto di gioco in cui.....

Capitolo 5

Conclusioni e lavori futuri

Aggiungere un ulteriore livello di learning sopra il meta-learner, che apprende dal supervisore qual è l'intervallo di tempo migliore in cui cambiare AI scelta.

Bibliografia

- [1] Lawrence, R. (1997). Using neural networks to forecast stock market prices. University of Manitoba, 333, 2006-2013.
- [2] White, H. (1988). Economic prediction using neural networks: The case of IBM daily stock returns.
- [3] Phua, P. K. H., Ming, D., & Lin, W. (2000, July). Neural network with genetic algorithms for stocks prediction. In Fifth Conference of the Association of Asian-Pacific Operations Research Societies, 5th-7th July, Singapore. sn.
- [4] Kim, K. J., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert systems with Applications*, 19(2), 125-132.
- [5] Garliauskas, A. (1999, October). Neural network chaos and computational algorithms of forecast in finance. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028) (Vol. 2, pp. 638-643)*. IEEE.
- [6] Schumann, M., & Lohrbach, T. (1993, January). Comparing artificial neural networks with statistical methods within the field of stock market prediction. In [1993] *Proceedings of the Twenty-sixth Hawaii International Conference on System Sciences (Vol. 4, pp. 597-606)*. IEEE.
- [7] Yoon, Y., Swales Jr, G., & Margavio, T. M. (1993). A comparison of discriminant analysis versus artificial neural networks. *Journal of the Operational Research Society*, 44(1), 51-60.
- [8] Kim, J. W., Weistroffer, H. R., & Redmond, R. T. (1993). Expert systems for bond rating: a comparative analysis of statistical, rule-based and neural network systems. *Expert systems*, 10(3), 167-172.
- [9] Patuwo, E., Hu, M. Y., & Hung, M. S. (1993). Two-group classification using neural networks. *Decision Sciences*, 24(4), 825-845.

- [10] Subramanian, V., Hung, M. S., & Hu, M. Y. (1993). An experimental evaluation of neural networks for classification. *Computers & operations research*, 20(7), 769-782.
- [11] Chalup, S., & Maire, F. (1999, July). A study on hill climbing algorithms for neural network training. In *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99* (Cat. No. 99TH8406) (Vol. 3, pp. 2014-2021). IEEE.
- [12] Hong, T., & Han, I. (2002). Knowledge-based data mining of news information on the Internet using cognitive maps and neural networks. *Expert systems with applications*, 23(1), 1-8.
- [13] Fung, G. P. C., Yu, J. X., & Lam, W. (2002, May). News sensitive stock trend prediction. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 481-493). Springer, Berlin, Heidelberg.
- [14] Kohara, K., Ishikawa, T., Fukuhara, Y., & Nakamura, Y. (1997). Stock price prediction using prior knowledge and neural networks. *Intelligent Systems in Accounting, Finance & Management*, 6(1), 11-22.
- [15] Fung, G. P. C., Yu, J. X., & Lam, W. (2003, March). Stock prediction: Integrating text mining approach using real-time news. In *2003 IEEE International Conference on Computational Intelligence for Financial Engineering, 2003. Proceedings.* (pp. 395-402). IEEE.
- [16] Kuo, R. J., Chen, C. H., & Hwang, Y. C. (2001). An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. *Fuzzy sets and systems*, 118(1), 21-45.
- [17] Kuo, R. J., Lee, L. C., & Lee, C. F. (1996, October). Integration of artificial neural networks and fuzzy delphi for stock market forecasting. In *1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems* (Cat. No. 96CH35929) (Vol. 2, pp. 1073-1078). IEEE.
- [18] Lowe, D., & Webb, A. R. (1991, February). Time series prediction by adaptive networks: A dynamical systems perspective. In *IEE Proceedings F (Radar and Signal Processing)* (Vol. 138, No. 1, pp. 17-24). IET Digital Library.
- [19] Maddala, G. S., & Lahiri, K. (1992). *Introduction to econometrics* (Vol. 2). New York: Macmillan.
- [20] Hellstrom, T., & Holmstrom, K. (1998). Predicting the stock market. Unpublished Thesis, Malardalen University, Department of Mathematics and Physics, Vasteras, Sweden.

- [21] Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417.
- [22] Tsibouris, G., & Zeidenberg, M. (1995). Testing the efficient markets hypothesis with gradient descent algorithms. In *Neural networks in the capital markets* (Vol. 8, pp. 127-136). Chichester, UK: Wiley.
- [23] Van Eyden, R. J. (1996). *The Application of Neural Networks in the Forecasting of Share Prices* (Finance and Technology Publishing, Haymarket, VA).
- [24] Lawrence, R. (1997). Using neural networks to forecast stock market prices. University of Manitoba, 333, 2006-2013.
- [25] Van Eyden, R. J. (1996). *The Application of Neural Networks in the Forecasting of Share Prices* (Finance and Technology Publishing, Haymarket, VA).
- [26] Refenes, A. P., Zapranis, A. D., & Francis, G. (1995). Modeling stock returns in the framework of APT: a comparative study with regression models. In *Neural networks in the capital markets* (Vol. 7, pp. 101-126). John Wiley & Sons Chichester.
- [27] Steiner, M. (1995). Neural networks as an alternative stock market model.
- [28] Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent data analysis*, 1(3), 131-156.
- [29] Hiemstra, Y. (1995). Modeling structured nonlinear knowledge to predict stock market returns. *Chaos & Nonlinear Dynamics in the Financial Markets: Theory, Evidence and Applications*, Irwin, Chicago, IL, 163-175.
- [30] Tsaih, R., Hsu, Y., & Lai, C. C. (1998). Forecasting S&P 500 stock index futures with a hybrid AI system. *Decision Support Systems*, 23(2), 161-174.
- [31] Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5), 988-999.
- [32] Yang, H., Chan, L., & King, I. (2002, August). Support vector machine regression for volatile stock market prediction. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 391-396). Springer, Berlin, Heidelberg.
- [33] Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5), 988-999.

- [34] Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1-2), 307-319.
- [35] Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *omega*, 29(4), 309-317.
- [36] Tay, F. E. H., & Cao, L. J. (2001). A comparative study of saliency analysis and genetic algorithm for feature selection in support vector machines. *Intelligent Data Analysis*, 5(3), 191-209.
- [37] Trippi, R. R., & Turban, E. (1992). *Neural networks in finance and investing: Using artificial intelligence to improve real world performance*. McGraw-Hill, Inc..
- [38] Kim, K. J. (2004). Toward global optimization of case-based reasoning systems for financial forecasting. *Applied intelligence*, 21(3), 239-249.
- [39] Ng, A., & Fu, A. W. C. (2003, April). Mining frequent episodes for relating financial events and stock trends. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 27-39). Springer, Berlin, Heidelberg.
- [40] Maheu, J. M., & McCurdy, T. H. (2004). News arrival, jump dynamics, and volatility components for individual stock returns. *The Journal of Finance*, 59(2), 755-793.
- [41] Fawcett, T., & Provost, F. J. (1996, August). Combining Data Mining and Machine Learning for Effective User Profiling. In *KDD* (pp. 8-13).
- [42] Hong, T., & Han, I. (2004). Integrated approach of cognitive maps and neural networks using qualitative information on the World Wide Web: the KBNMiner. *Expert Systems*, 21(5), 243-252.
- [43] Schmelzer, R. (2020, January 3). How Do You Test AI Systems?. *Forbes*. Last accessed 21th Feb 2020: <https://www.forbes.com/sites/cognitiveworld/2020/01/03/how-do-you-test-ai-systems>.
- [44] Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning* (Vol. 135). Cambridge: MIT press.
- [45] Guijarro-Berdiñas, B., & Alonso-Betanzos, A. (2002). Empirical evaluation of a hybrid intelligent monitoring system using different measures of effectiveness. *Artificial Intelligence in Medicine*, 24(1), 71-96.