Machine Learning for Finance

Applicazioni in Credit Scoring, Real Estate e Trading Algoritmico

Edoardo Vittori, Cristian Giacomini, Emilio Colombo

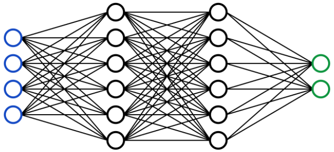
EMF Campus
9 Novembre 2022

AGENDA

Intelligenza Artificiale e Machine Learning

Descrizione ed esempi

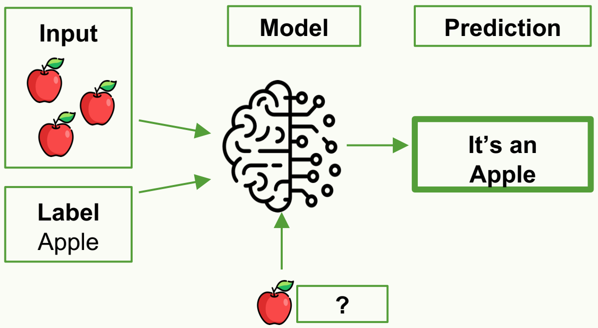
	Descrizione	Esempi
Artificial Intelligence	• Un **programma** che può **percepire**, **ragionare** ed **agire**	• Robotica • Guida autonoma • Virtual Assistants
Machine Learning	• **Algoritmi** che, senza avere istruzioni dirette, **imparano** e **migliorano** con l'**esposizione** ai **dati**	• Pricing Automatico • Predictive maintenance • Credit Rating • Robot Advisors • Stock Picking • Trading • Natural Language Processing • Image Recognition



Machine Learning

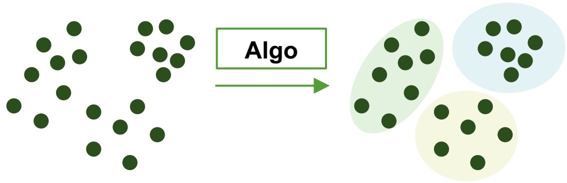
Principali paradigmi del Machine Learning

Supervised Learning



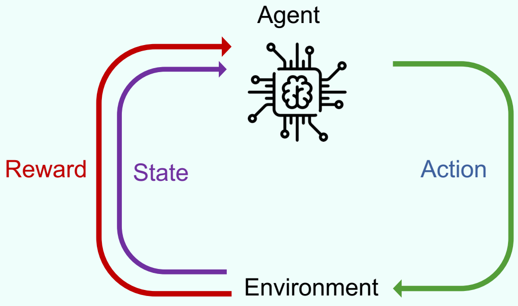
Utilizzo
- Regression
- Classification

Unsupervised Learning



Utilizzo
- Clustering
- Dimensionality reduction
- Feature extraction

Reinforcement Learning

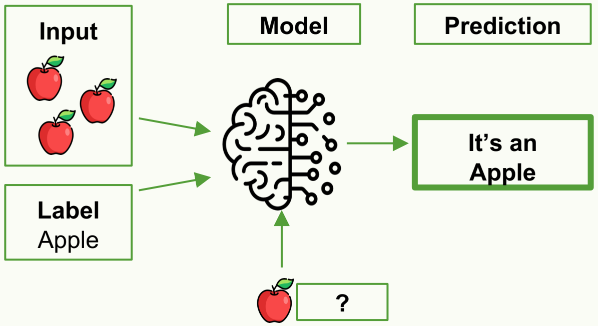


Utilizzo
- Optimal Control
- Game playing

Machine Learning per la Finanza

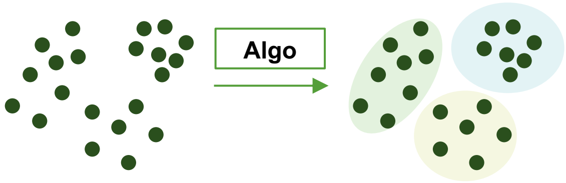
Principali paradigmi del Machine Learning

Supervised Learning



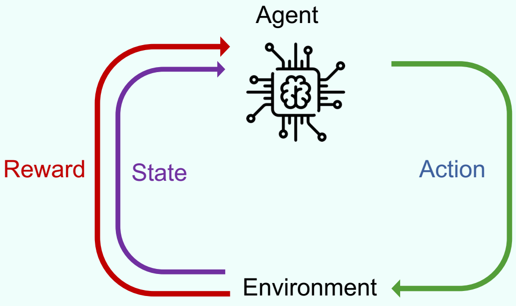
Utilizzo
- Price prediction
- Automatic Valuation Models
- Credit scoring

Unsupervised Learning



Utilizzo
- Regime Detection
- Client Clustering
- Fraud detection

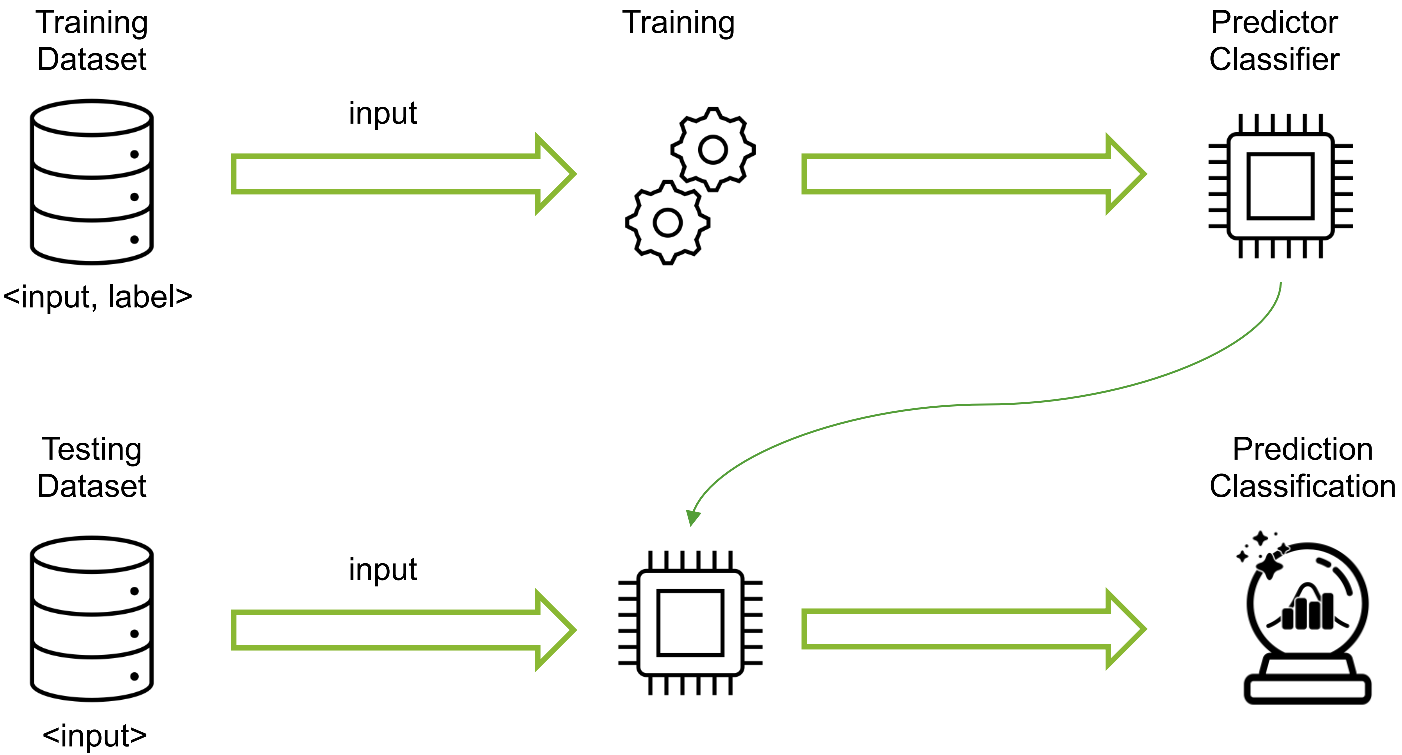
Reinforcement Learning



Utilizzo
- Trading
- Portfolio optimization
- Market Making
- Hedging
- Optimal Execution

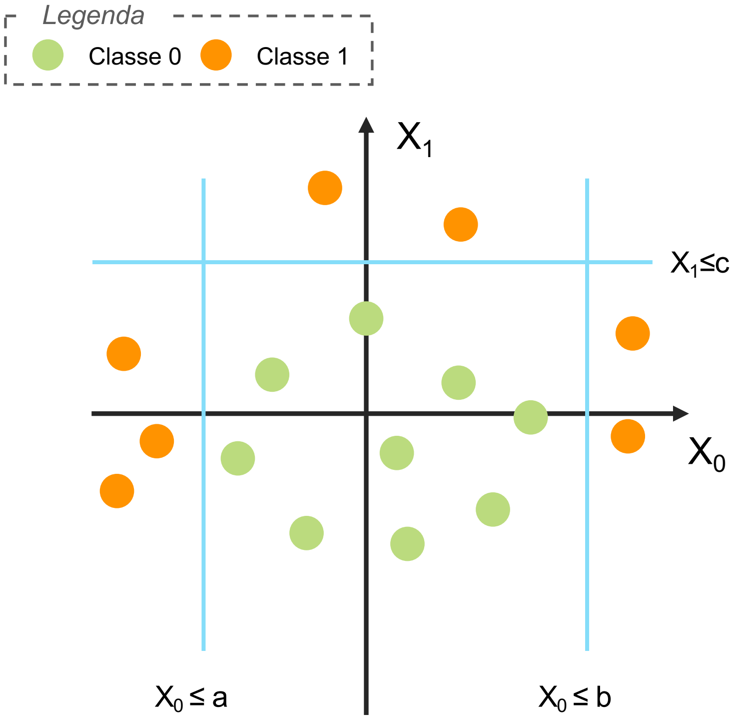
Supervised Learning

Training and testing

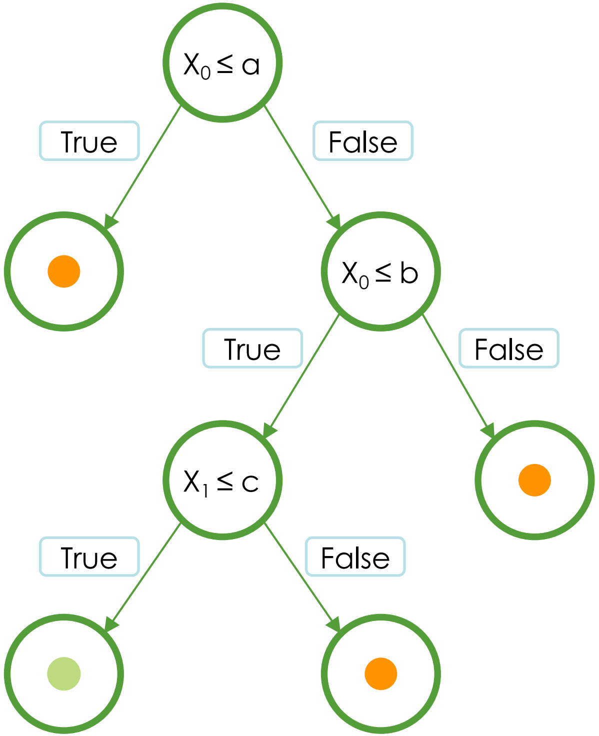


Decision Trees

Esempio dati training



Esempio albero



Caratteristiche

Utilizzo

- Classificazione
- Regressione

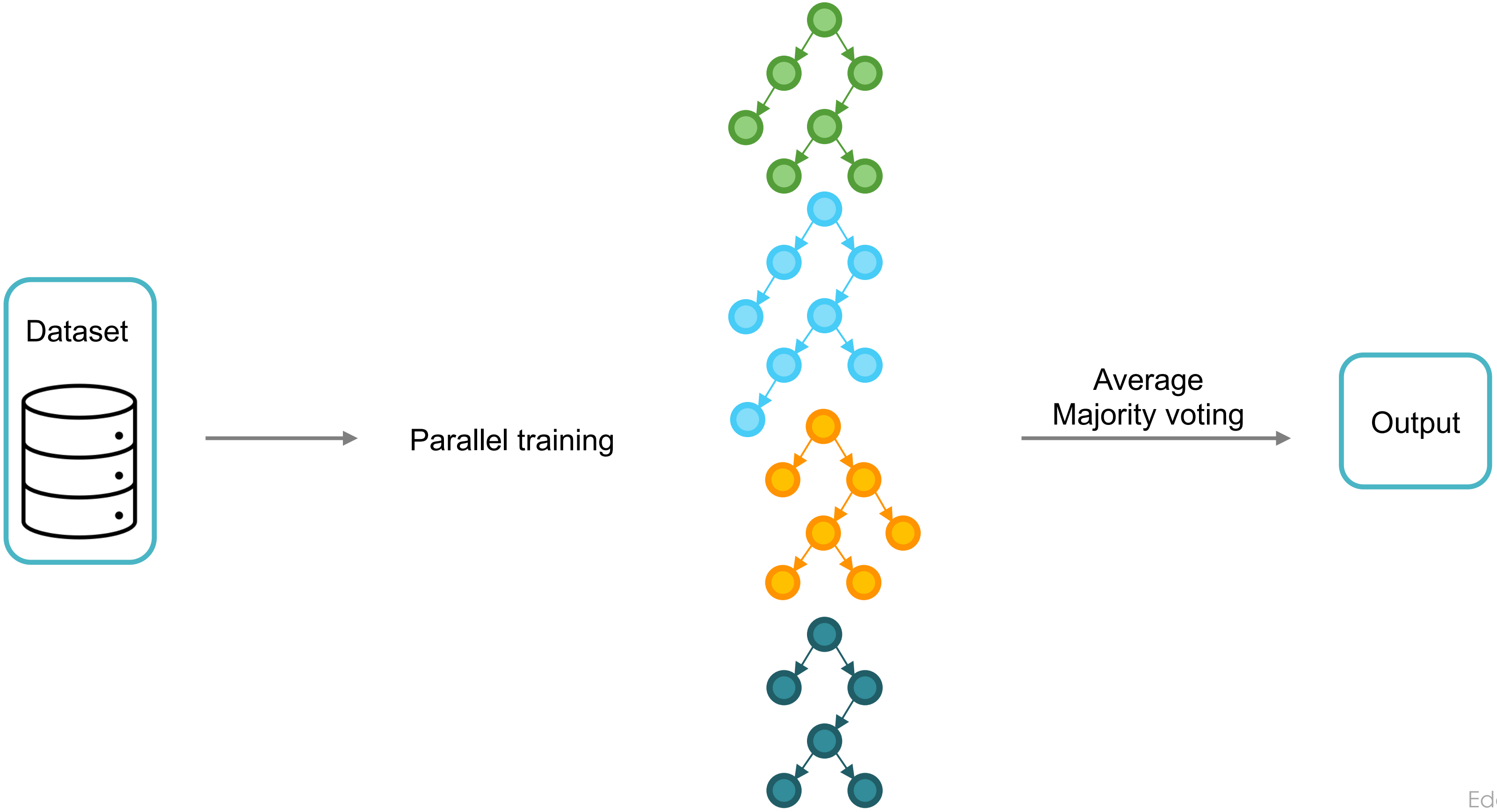
Splitting Criterion

- Gini impurity
- Entropy
- Information Gain
- Max depth

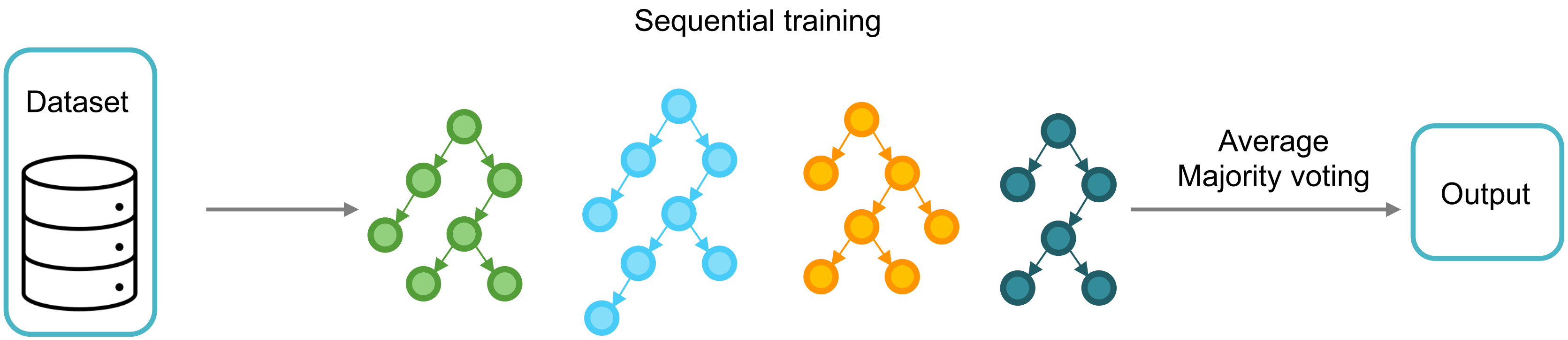
Further developments

- Random forests
- XGBoost

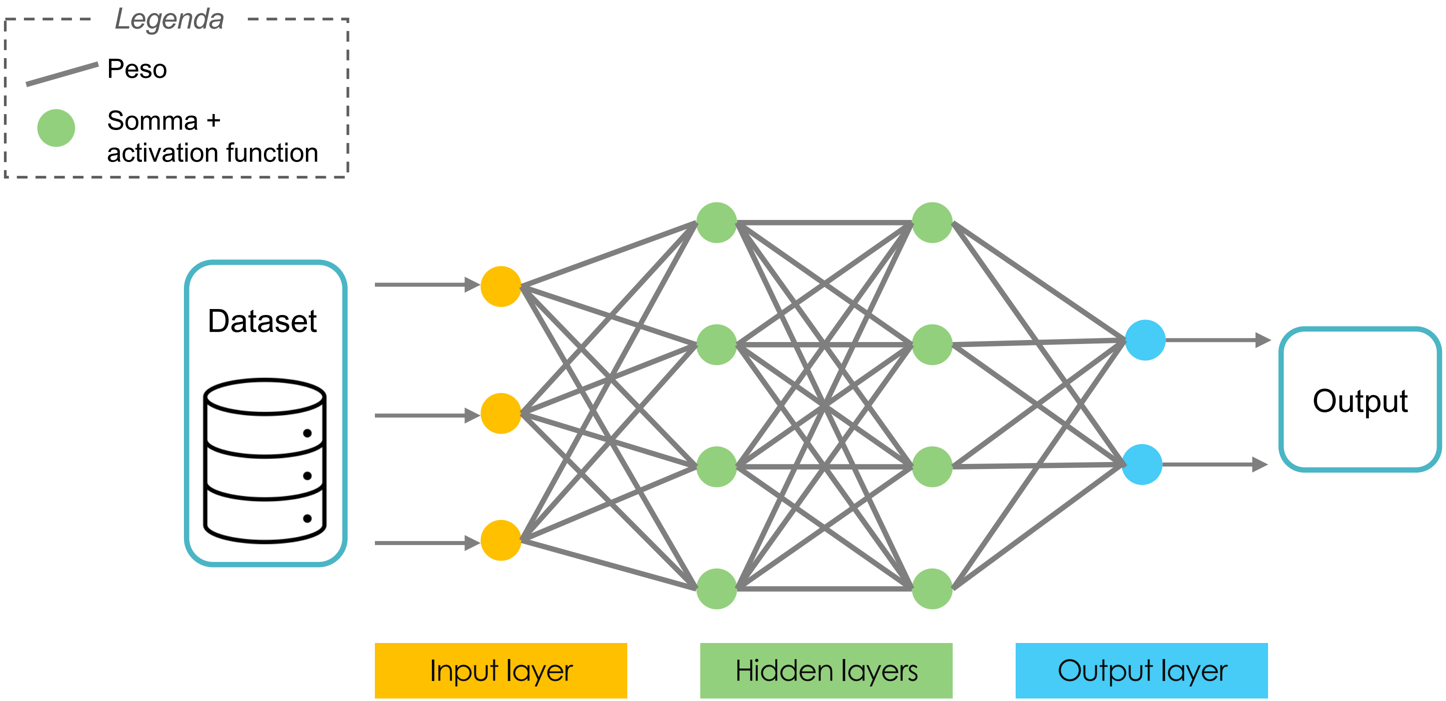
Random Forests



eXtreme Gradient Boosting (XGBoost)



Neural Networks



Caratteristiche

Utilizzo

- Classificazione
- Regressione

Training

- Back-propagation

Parametri

- Loss
- Activation function
- Batch size
- Epochs

AGENDA

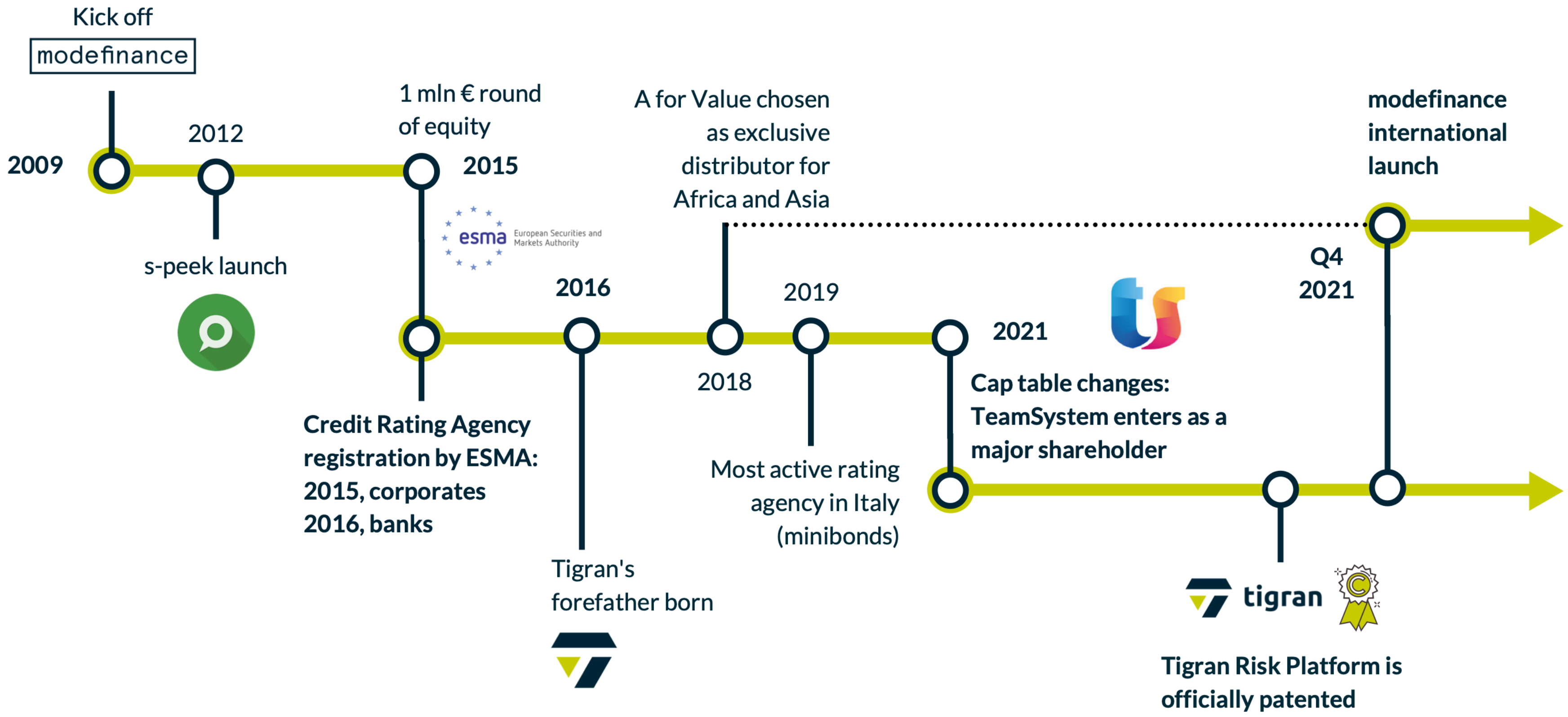
1 Introduzione al Machine Learning

2 Use cases
- **Credit Risk Management**
- Automated Valuation Models
- Algorithmic trading

Data Science and Machine Learning in Credit Risk Management

Cristian Giacomini - CEO Modefinance International

roadmap


Kick off
modefinance
2009
2012
s-peek launch
1 mln € round of equity
2015
esma European Securities and Markets Authority
Credit Rating Agency registration by ESMA: 2015, corporates 2016, banks
2016
Tigran's forefather born
A for Value chosen as exclusive distributor for Africa and Asia
2018
2019
Most active rating agency in Italy (minibonds)
2021
Cap table changes: TeamSystem enters as a major shareholder
tigran
Tigran Risk Platform is officially patented
Q4 2021
modefinance international launch

EU Fintech credit rating agency CRA & ECAI

esma European Securities and Markets Authority

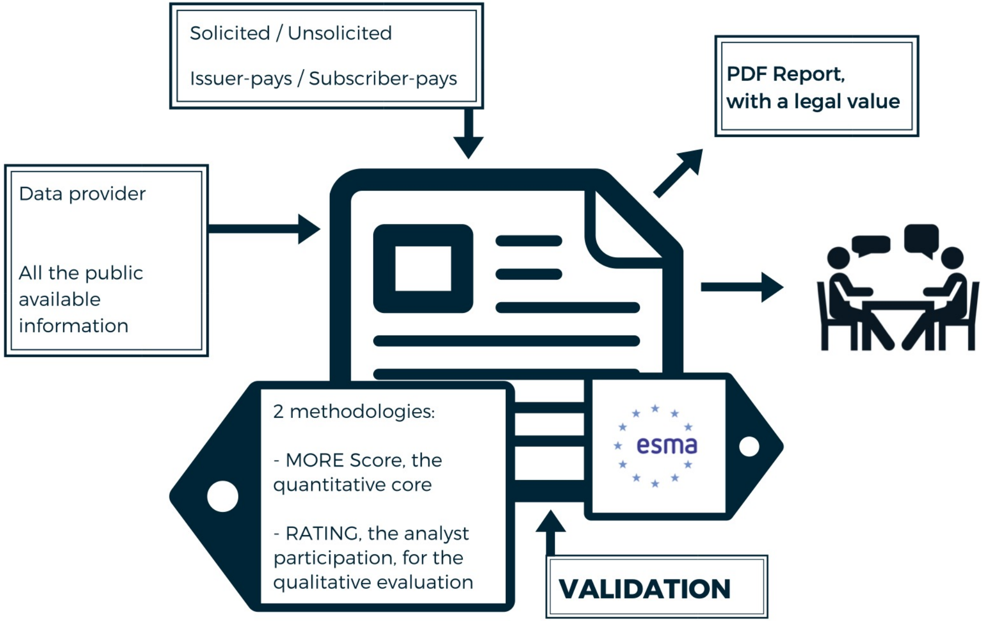
Date: 10 July 2015
ESMA/2015/1127

PUBLIC STATEMENT

ESMA registers modeFinance as credit rating agency

The European Securities and Markets Authority (ESMA) has formally approved the registration of modeFinance S.r.l., based in Trieste, Italy, as a credit rating agency (CRA) under Article 16 of the CRA Regulation. The registration takes effect from 10 July 2015.

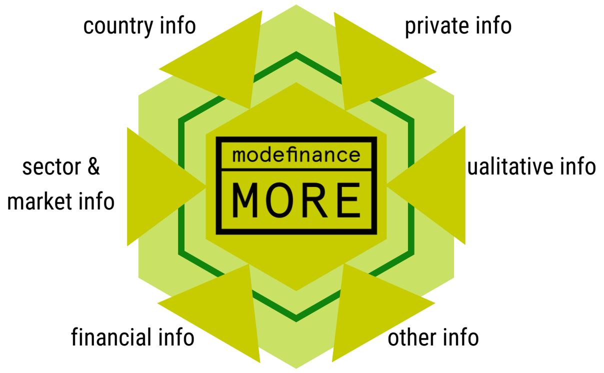
modeFinance's registration as a CRA means that its credit ratings can be used for regulatory purposes within the European Union.



modefinance 2022 All information shall be treated as strictly confidential.

modefinance technology core

modefinance develops custom and tailor-made models constantly, since the beginning in 2009.
A set of proprietary big data methodologies and machine learning algorithms, MORE, evaluates any company and financial institution in the world, without geographic, sector or country limits.



A broad set of models related to credit risk management is available, among these:

- Credit scoring models
- Credit limit models
- Default prediction models
- Pricing models
- Credit decision engines
- Many others, on premise

modefinance 2022 All information shall be treated as strictly confidential.

use of AI within modefinance models

In general, there are 2 ways according to which modefinance employs AI for the models' development:

- Feature selection and data clustering for model development;
- As a component used to enhance the performance of existing models.

Currently AI is not used as an "end to end solution". This is mostly to ensure our clients the maximum transparency and control over the models modefinance deploys, to face customer needs.

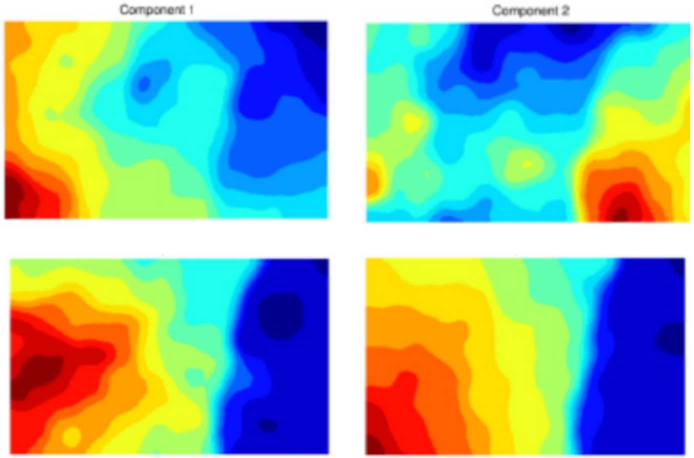
Differently, for custom projects we are able to respond to customers' requests when specifically asked for AI-based models.

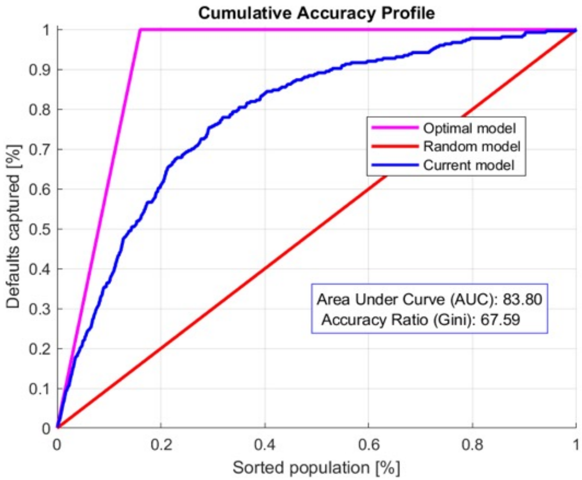
In the next few slides you will find some examples of these two events.

modefinance 2022 All information shall be treated as strictly confidential.

AI for models' development

modefinance adopted and employs AI technology daily since the very beginning for the purpose of exploring data sources and integration in the process of developing new models. Even for the development of modefinance's core model, the MORE score (about 2008-2009), we deepened the analysis and use of Neural Networks (in the form of SOM) to analyze the discriminating capacity of variables, derived from the companies' financial accounts, with respect to the default event.





modefinance 2022 All information shall be treated as strictly confidential.

AI for models' development

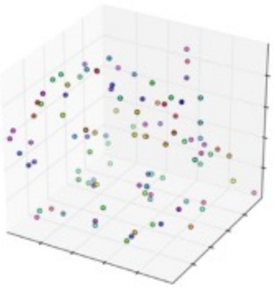
	31/12/2021 Excel UNCONSOLIDATED	31/12/2020 Excel UNCONSOLIDATED
Turnover (€)	253,144,344	209,166,878
modefinance score	BB	B
One year probability of default	2.44%	2.88%
Confidence	100.00%	100.00%
Solvency ratios		
Leverage ratio	2.93	2.70
Financial leverage	1.73	1.75
Total asset/Total liabilities	1.34	1.37
Liquidity ratios		
Current ratio	0.90	0.85
Quick ratio	0.90	0.85
Cash cycle ratio	17.00	14.00
Profitability ratios		
Return on investement ROI	5.27%	3.33%
Return on equity ROE	8.34%	11.79%
Asset turnover	1.37	1.32

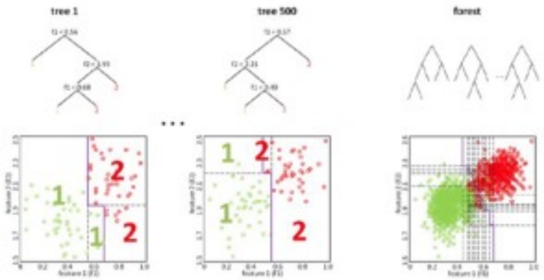
Today with MORE score modefinance evaluates more than 25 million companies per year, all around the world.

modefinance 2022 All information shall be treated as strictly confidential.

AI for models' development

Recently, modefinance developed a model to estimate the impact of the Covid-19 effect on the creditworthiness of companies (2019). The model has foreseen the use of Random Forests to identify, among several different features, those best describing the evolution of the companies' creditworthiness.





modefinance 2022 All information shall be treated as strictly confidential.

AI for models' development

This models also factors macro-economic and sector data, to provide a real-time-forecast of the most probable evolution of the companies' creditworthiness.

Via such model today modefinance provides forecasts on the evolution of about 18 million companies worldwide, on a monthly basis.

	31/12/2021 UNCONSOLIDATED	31/12/2022
modefinance score	BB	BB
One year probability of default	2.44%	1.70%
Solvency	BBB	B
Profitability	BB	
Liquidity	B	
Coverage	B	
Months	12	

Balance sheet (€)		
+ Total assets	178,019,195	182,533,599
+ Shareholders funds	45,248,529	50,364,876
+ Total liabilities	132,770,666	132,168,723
Total shareh. funds & liab.	178,019,195	182,533,599
Short term debts	73,104,853	68,502,567
Long term debt	4,596,944	4,264,515
Cash & Cash equivalent	3,004,563	11,043,476

modefinance 2022 All information shall be treated as strictly confidential.

AI for models' development

Modefinance has been involved into several custom projects, where AI is a component of tailored models:

- One of these, a project aimed to reconcile bank transaction data (derived from the open banking) with the invoices received or issued by a company.
- For a similar project, we did reconstruct the supply chain of a company, starting from open banking data.

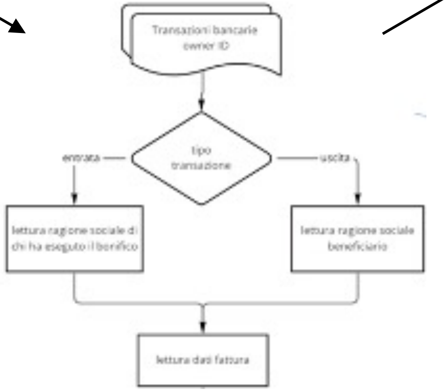
For both this projects modefinance has employed Natural Language Processing (NLP) algorithms applied to the entity recognition: from transactional and invoice data we extrapolated all the text containing information relative to the companies' related to each transaction.

Doing this, we are able to both identify those companies, AND to evaluate the degree of riskiness of the cash flows of a company.

modefinance 2022 All information shall be treated as strictly confidential.

AI for models' development

```
{
    "owner_id": "05377384-f18e-48c7-8dfa-4b9f994059a4",
    "transactions": [{
            "description": "Commissioni COMM/SP ADUE Mo
            "id": "60793ca785044f40698a0fd6",
            "amount": 800,
            "cbi_reason": 16
        }, {
            "description": "Commissioni COMM/SP ADUE Mo
            "id": "60793ca785044f40698a0fd6",
            "amount": 800,
            "cbi_reason": 16
        }
    ]
}
```

Thousands of customers are using this algorithm to reconcile their invoices with bank transactions and to monitor their supply chain.

modefinance 2022 All information shall be treated as strictly confidential.

custom models

Other projects see modefinance involved as a TechFin partner: for example, a project where the customer wants to explore the application of AI to enhance the prediction of a very particular (non disclosable) credit event.

The customer previously implemented its own internal model for the prediction of the said credit event, and we have been involved to solve the task of improving that existing model, and to see if the application of AI technologies could further enhance the predictive capacity.

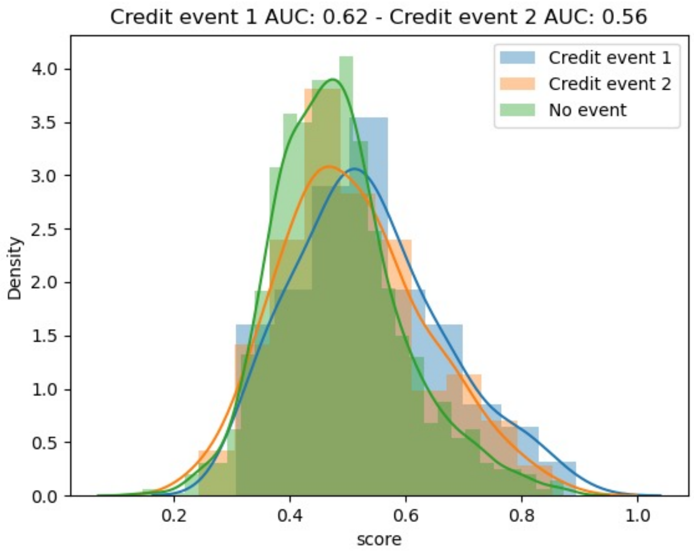
The results achieved so far are quite satisfactory: the application of XGboost has improved the predictive capacity obtained with the "canonical" model of several percentage points.

modefinance 2022 All information shall be treated as strictly confidential.

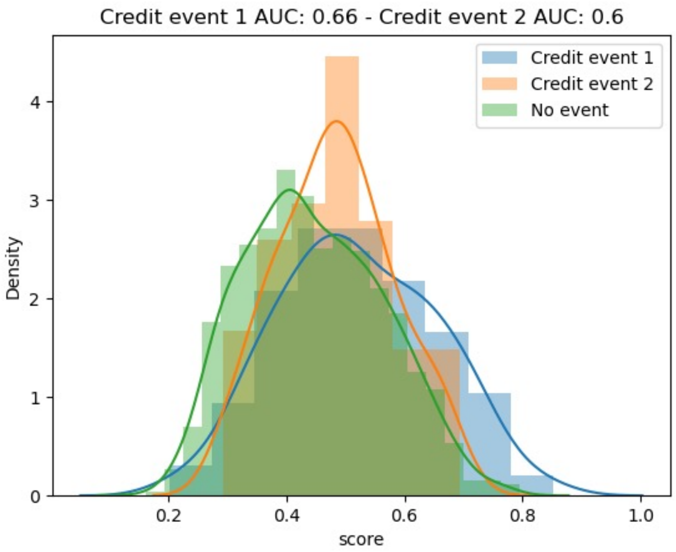
custom models

An example of XGboost application:

Canonical model used to predict the credit event



XGboost applied to predict the credit event



modefinance 2022 All information shall be treated as strictly confidential.

trusted by

بيان للمعلومات الائتمانية
Bayan Credit Bureau

GENERALI INVESTMENTS

SAMSUNG

ETIHAD CREDIT INSURANCE ECI الاتحاد لائتمان الصادرات

TÜRK EXIMBANK

DUPONT

eni

saipem

PIAGGIO

Mercedes-Benz

FIDELITY BANK
Believe with us

BANCA PROGETTO

Metsä

MEDIOCREDITO CENTRALE

INVITALIA

Lindt
MASTER CHOCOLATIER SINCE 1845

Barilla

مضاربة
mudaraba

modefinance evaluations are distributed by

BUREAU VAN DIJK
A Moody's Analytics Company

AGENDA

1 Introduzione al Machine Learning

2 Use cases
- Credit Risk Management
- **Automated Valuation Models**
- Algorithmic trading

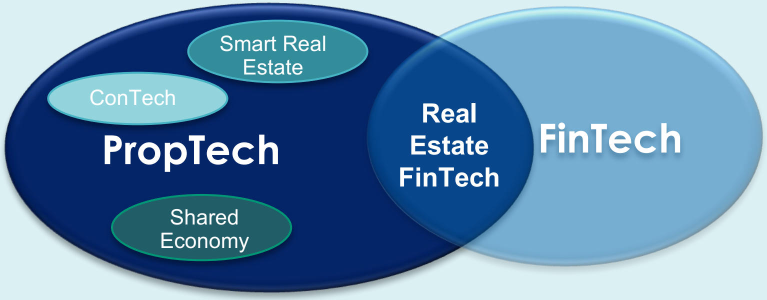
Real Estate and Artificial Intelligence Automated Valuation Models

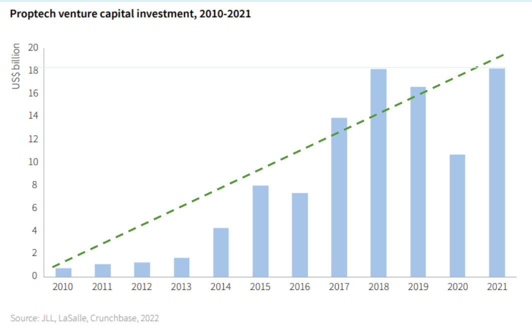
Emilio Colombo - Cerved

Real Estate & Artificial Intelligence

PropTech Oveview

L'**AI** è entrata formalmente nel mondo del Real Estate con l'avvento del **PropTech (Property+Technology)**, definito come parte della **trasformazione digitale del settore immobiliare**, mediante **soluzioni, tecnologie e strumenti per l'innovazione dei processi, prodotti e servizi** che lo caratterizzano.

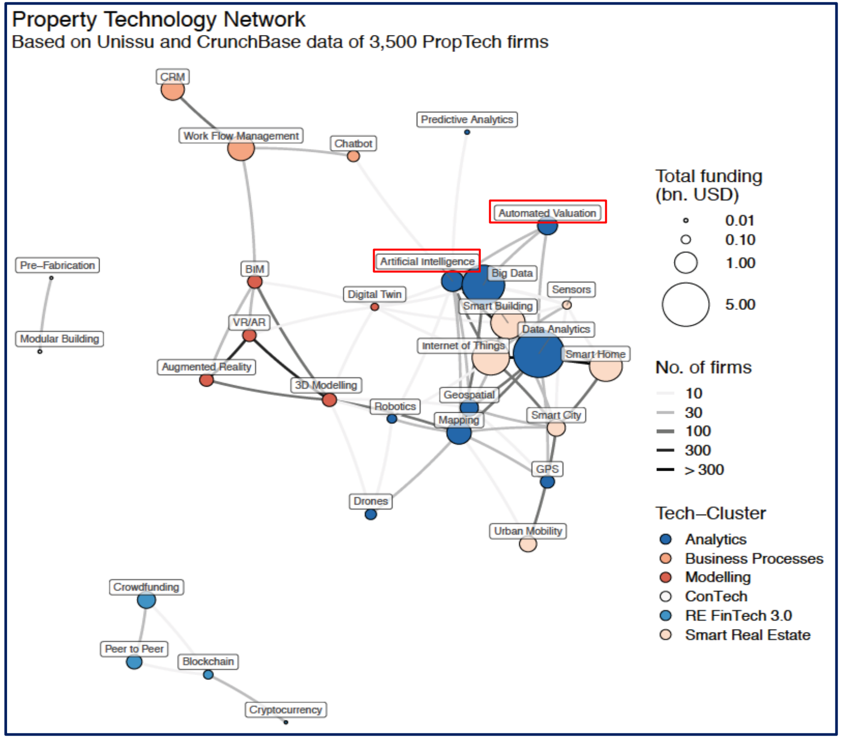




Con un **CAGR** del **16,8%**, il **mercato del Proptech** salirà dai **US$ 18,2 Bn** del **2022** agli **86,5 Bn** del **2036**.

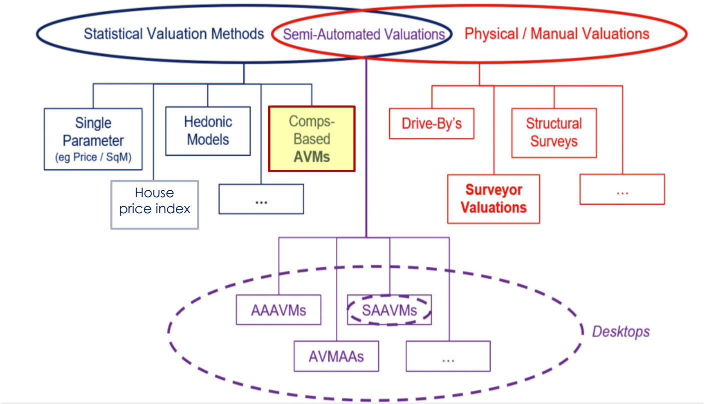
I mercati di USA e UK, già sviluppati, sfrutteranno le sinergie dei mercati in forte crescita come **Cina, Giappone ed Europa** con rispettivi **CAGR** del **23,7%, 26,7% e 17,1%.**

Dati: FMI

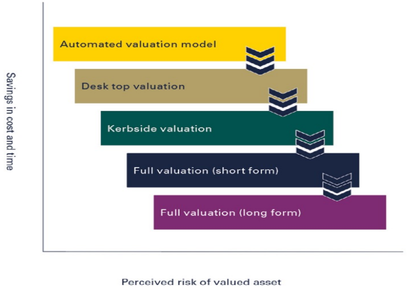


Automated Valuation Model

Il Machine Learning applicato al settore delle valutazioni immobiliari



Gli **AVM** sono definiti come "...una o più tecniche matematiche per fornire una **stima del valore** di un dato **immobile** ad una **certa data**, accompagnata da una **misura di confidenza nell'accuratezza** del risultato, **senza intervento umano** dopo la fase iniziale"[1]



Tali sistemi **permettono** anche grazie al ML di:

- **valutare massivamente migliaia di asset** in pochissimi minuti in funzione della rischiosità
- **ridurre** sensibilmente **i costi per le valutazioni preliminari** di grandi portafogli immobiliari
- dare una **misura di accuratezza della valutazione statistica** in funzione dei diversi scenari

1) Information paper di RICS (Royal Institution of Chartered Surveyors
2) European Mortgage Federation
3) European AVM Alliance
4) Guidelines on Loan Origination and Monitoring

1960 - 1990
AVMs have their origins in **North America** (1960), the **first commercial application** was created in 1981 and **began** to be **developed** in the **UK** in the 1990s.

2008 - 2009
Robinson & Dawnie demonstrated importance of AVM. In 2009, **EMF**[2]: 'AVM is a **useful** and efficient tool **when used** appropriately **by an experienced operator**'.

2010
American Bankers Association [2010]: 'Institutions may **employ AVM**s for a variety of uses, such **as loan underwriting** and **portfolio monitoring**'.

2013 - 2018
2013 - **EU Parliament**: Institutions may use **statistical methods to monitor** property's **values** and to plan revaluation. 2018: **Standards** for statistical valuation methods for residential properties **in Europe** – EAA[3]

2021
EBA Guidelines LOM[4] define **advanced statistical models** as suitable for **use in origination**, **monitoring** and **revaluation** in addition to expert **back-testing**

AVM: Esempio di valutazione

Dagli input agli output

Your property summary

This exclusive report provides a unique up-to-date insight into the value of your property and analysis of the local market, for **2 Copper Row London, SE1 2LH**

Property type **Purpose Built Flat**	Total floor area **753 sqft**	Year built **1900**
Receptions **1**	Bedrooms **2**	Bathrooms **1**

Market insights

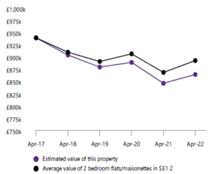
Southwark

95%

Properties near you are achieving an average of **95%** of the asking price

21 weeks on market

Properties near you spend an average time on the market of **21 weeks**



Price change over the last 5 years of 2 bedroom flats/maisonettes in **SE1 2**

No search data

Recent activity near you

Horselydown Lane
Distance: 0.02 miles
Type: 2 bedroom purpose built flat
Size: 1119 sqft
For sale price: £1,450,000
Last sold price: £1,202,295
Sold date: June 2017
Status: Under offer

Gainsford Street
Distance: 0.04 miles
Type: 3 bedroom flat/maisonette
Size: 1334 sqft
Last sold price: £1,200,000
Sold date: September 2021

Your valuation powered by Hometrack

Estimated capital value **£865,000**	Estimated value range **£780,000 to £950,000**	Confidence in valuation estimate **medium**
Last sale **unknown**	Estimated rental value **£2,430 pcm**	
Estimated gross rental yield **3.4%**		

INPUT — **OUTPUT**

Dati immobile

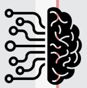
- Localizzazione
- Tipologia di immobile
- Consistenza (mq/vani)
- Anno/Periodo di costruzione
- Piano/Ascensore
- Esposizione, finiture, etc..

Dati mercato/territoriali/economici

- Serie storiche mercato immobiliare (prezzi, sconti, tempi, yield, etc.)
- Dati macroeconomici territoriali (PIL, disoccupazione, mutui, etc.)
- Dati socio demografici (Pol, popolazione, pendolarismo, etc.)
- Dati ambientali/Score settoriali

Dati immobili comparabili

- DB transazioni
- DB perizie fisiche
- Asking price
- DB CTU
- DB aggiudicazioni asta



Black Box

Valori statistici

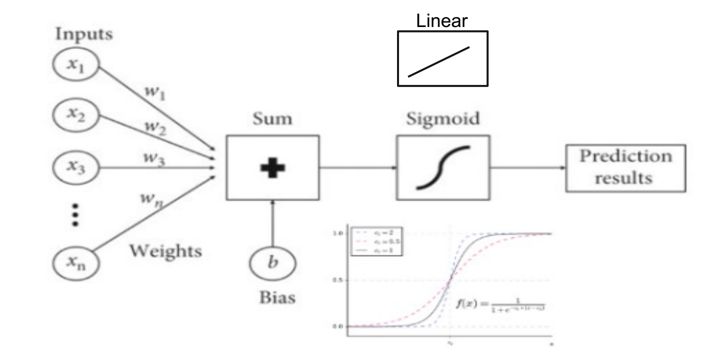
- Valore di mercato max, min, mid
- Valori prospettici (12, 24, 36 m)
- Confidenza della stima
- Valore di locazione/rendimento
- Valore cauzionale

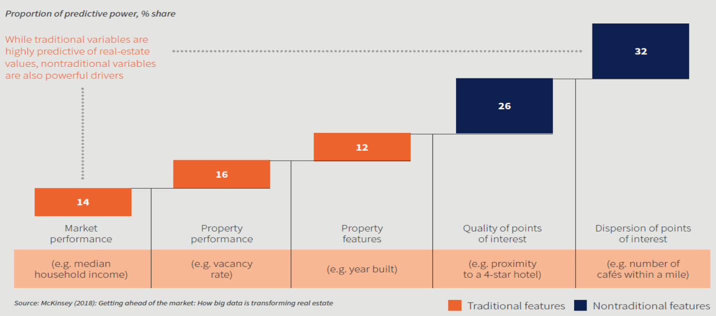
AVM: Come funzionano

Hedonic Approach AVM vs Machine Learning Approach AVM

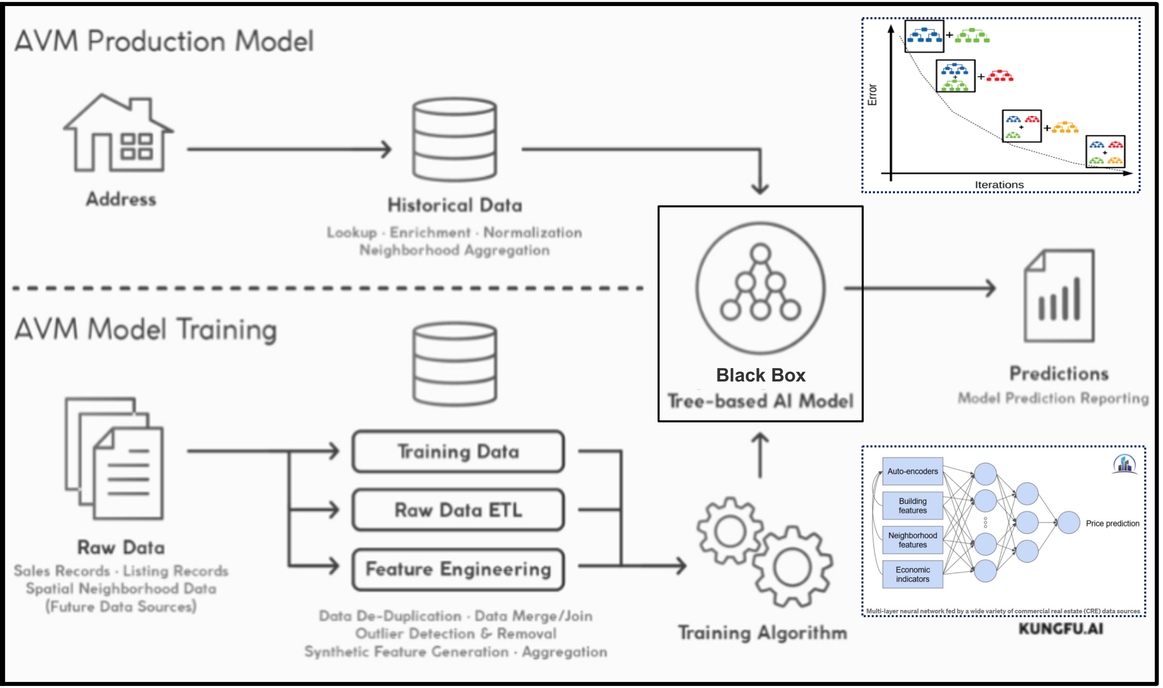
Tutti i **modelli** vengono **valutati secondo due capacità**: **la capacità inferenziale** e la **capacità predittiva**. La prima consiste nell'abilità del modello nell'**individuare rapporti causa-effetto tra le variabili** spiegate e le variabili indipendenti, tipica dei **modelli edonici**.
La seconda risiede nella capacità del modello di elaborare risultati di output corrispondenti al valore dei dati reali.

Approccio Edonico





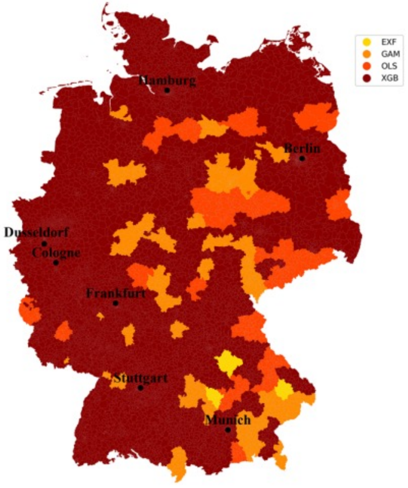
Machine Learning



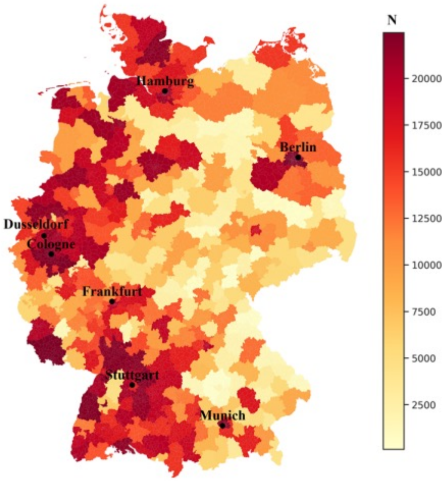
AVM: Case study in Germania

Performance a confronto dei principali modelli AVM

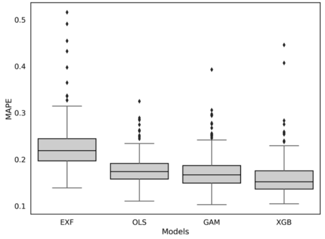
Secondo uno recente studio[1] sull'automazione delle perizie in **Germania**, sono stati messi a confronto i 4 modelli di **AVM** più diffusi su un dataset di 1,2 M di market values di residenze. Il modello di **ML Extreme Grandient Boost** ha mostrato le **migliori performance predittive**.

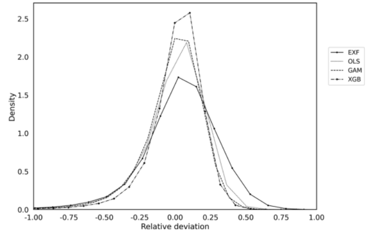


Best performing model of every adminstrative district

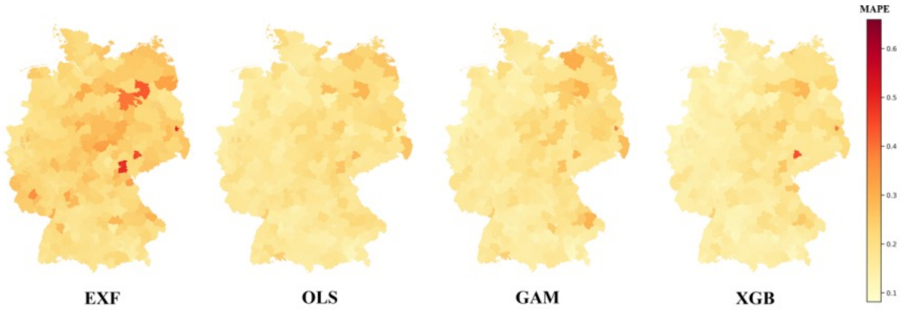


Number of observations per district





Models	MAPE	MdAPE	PE(10)	PE(20)	R^2
EXF	0.2130	0.1624	0.3267	0.5872	0.7735
OLS	0.1736	0.1311	0.3937	0.6940	0.8654
GAM	0.1646	0.1202	0.4273	0.7276	0.8664
XGB	0.1465	0.1084	0.4665	0.7786	0.8995
Error	Mean Absolute Percentage Error	Median Absolute Percentage Error	Error bucket (Dev<10%)	Error bucket (Dev<20%)	Coefficient of determination



Expert Function: similarity function

Ordinary Least Squares: hedonic price fuction

Generalized Additive Models: hedonic price function

eXtreme Gradient Boost: machine laerning

1) From human business to machine learning—methods for automating real estate appraisals and their practical Implications - Moritz Stang · Bastian Krämer · Cathrine Nagl ·Wolfgang Schäfers

AVM: I Principali Utilizzi

Mercato e Needs

Market	Needs - Conventional Applications	Future needs - Innovative Applications
Banks	▪ Migliorare tempi/costi di erogazioni ipotecarie ▪ Ridurre drop-out di clienti dopo domanda di mutuo ▪ Monitoraggio continuo del valore delle garanzie immobiliari ▪ Valutazione preliminari per cartolarizzazioni	▪ Previsione del rischio immobiliare di mercato con riduzione degli RWA ed accantonamenti ▪ Introduzione dei sistemi di allerta immobiliari ▪ Riduzione masse NPE ed impatto sociale
Italia	▪ Aggiornamento rendite catastali ▪ Monitoraggio tasse ed imposte immobiliari ▪ Individuazione di attività fraudolente	▪ Individuare aree di potenziale sostegno economico/sviluppo ▪ Individuare immobili per social/student housing ▪ Avviare progetti di crowdfunding/sviluppo sociale partecipato
Investors	▪ Analisi preliminari di portafogli ▪ Monitoraggio del valore immobiliare di portafoglio ▪ Scouting opportunità di investimento	▪ Individuare aree di potenziale sviluppo opportunistico ▪ Diffondere Home Equity ▪ Trading su alternative RE asset class
i-buyers	▪ Aumentare # clienti che impiegano la piattaforma ▪ Velocizzare il processo di vendita ▪ Migliorare la trasparenza della proposta di vendita	▪ Digitalizzare il processo di vendita immobiliare ▪ Offrire strumenti evoluti anche ad agenti immobiliari

AVM: I principali Market Players

Overview internazionale

Zillow Zestimates

AVM residenziale **leader in USA**. Realizza **110 M di valutazioni automatiche**. Basato su dati pubblici e inseriti dagli user sul portale di vendita Zillow. "*It is not an appraisal and can't be used in place of an appraisal*".

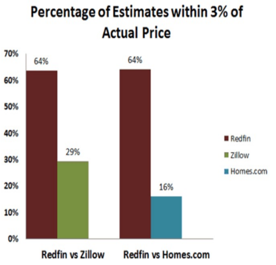
A livello nazionale **l'errore dalla corrente mediana è 3,2%** per le **residenze in vendita** e del **7,52%** per quelle **off markets.**

Metropolitan Areas	Median Error	Homes With Zestimates	Within 5% of Sales Price	Within 10% of Sales Price	Within 20% of Sales Price
Atlanta	3.1%	27.1K	70.1%	92.0%	98.5%
Austin	4.3%	12.8K	56.8%	87.2%	98.6%
Baltimore	3.1%	10.9K	70.5%	90.9%	96.3%
Boston	4.3%	14.4K	56.8%	86.4%	98.8%
Charlotte	3.3%	10.6K	67.5%	91.8%	98.7%

REDFIN

AVM **competitor americano** con un **errore** dalla corrente mediana del **2,23% per le residenze in vendita** e del **7,42%** per quelle **off markets.**

↑ Location	Estimate Count	% Median Error	Within 5%	Within 10%	Within 20%
Alabama	22,007	2.26%	73.15%	88.70%	96.17%
Arizona	51,555	1.92%	79.96%	93.17%	97.93%
Arkansas	13,235	2.36%	72.94%	89.14%	96.16%
California	134,120	2.17%	77.60%	92.45%	98.18%
Colorado	62,257	1.58%	84.57%	95.08%	98.59%
Connecticut	15,067	3.13%	67.41%	89.76%	98.18%
Delaware	6,064	1.90%	78.28%	91.76%	97.50%
District of Columbia	3,106	1.51%	84.52%	95.09%	99.22%
Florida	196,800	2.12%	77.28%	91.84%	97.74%
Georgia	83,979	2.00%	78.80%	92.27%	97.65%
All Locations	1,732,317	2.22%	75.89%	91.44%	97.65%



hometrack

AVM **leader UK** market. **17 delle 20 banche UK "top" usa AVM** come parte integrante dell'**erogazione di prestiti ipotecari**. Realizza **50 M valutazioni automatiche** ogni anno. Adottato in +50 Residential Mortgage-backed Securities (RMBS) e primo modello accreditato dalle principali agenize di Rating quali Moody's, Standard & Poor and Fitch.

EUROPEAN AVM ALLIANCE

CALCASA

Cerved Property Services

CRIF Together to the next level

Eiendomsverdi

hometrack

on-geo

tinsa

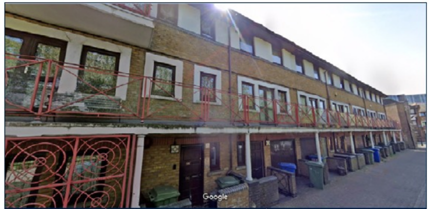
VärderingsData

AVM vs Appraisals

Vantaggi e Svantaggi

VANTAGGI	SVANTAGGI
Minore sforzo manuale da parte del valutatore	**Non** è dettagliata o **approfondita** come una valutazione umana.
Più economico rispetto alla valutazione tradizionale	**Non** può **tenere conto** di **alcuni fattori** come l'esposizione, le ristrutturazioni o il livello di manutenzione.
Valutazioni più rapide significano un'elaborazione più rapida dei prestiti	Le valutazioni possono essere basate su dati incompleti, imprecisi con basse frequenze. **(data lag bias)**
Meno soggetti a **errori e pregiudizi umani**	Può essere sbagliato di migliaia di dollari in mancanza di dati di vendita comparabili. **(data avaibility bias)**

Confronto fra performance di 3 AVM UK



2 Dec 2020	End terrace (west), northerly aspect faces other buildings	End terrace (east), northerly aspect faces Thames and City
JLL	£1,007,400 - £1,231,200 (£1,119,300)	£476,900 - £582,700 (£529,800)
Zoopla	£1.1 - £1.34 million (£1,220,000)	£1,150,000 - £1,410,000 (£1,280,000)
Your Move	£720,000 - £880,000 (£800,000)	£720,000 - £880,000 (£800,000)

3 Dec 2020	No particular note-worthy aspect features, but same floor plate as comparable	Rear of terrace is west-erly aspect over park, Thames and city
JLL	£660,600 - £807,400 (£734,000)	£660,600 - £807,400 (£734,000)
Zoopla	£610,000 - £915,000 (£763,000)	£1,000,000 - £1,110,000 (£1,060,000)
Your Move	£1,350,000 – £1,650,000 (£1,500,000)	£1,350,000 - £1,650,000 (£1,500,000)

Fonte: University of Oxford Research - the future of automatedreal estate valuations(AVMs)

Nuove prospettive per il Real Estate

AVM: Primo passo per una nuova era del Real Estate?

- **Predire** l'andamento del mercato residenziale
- Gestire e **prevenire** i **rischi immobiliari**
- **Rendere più trasparente** un mercato che soffre di opacità, lentezza e costi elevati nelle transazioni
- **Smaterializzare** un asset class che per definizione è **immobile** ed illiquida (tokenizzazione)
- **Attrarre nuovi investimenti** in un mercato che ha un valore stimato da Savils in circa 228 tr$

AGENDA

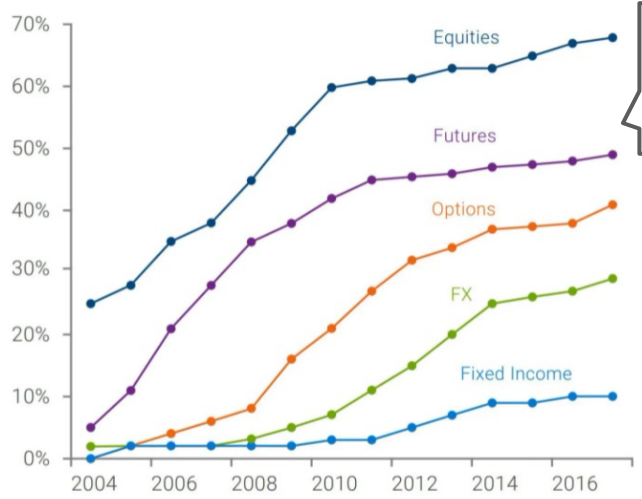
Algorithmic Trading

Edoardo Vittori – Intesa Sanpaolo

Trading Algoritmico

Mercato e tipologie di algoritmi nel trading

Quota di mercato degli algoritmi per asset class



As of 2017
Source: Goldman Sachs, Aite Group

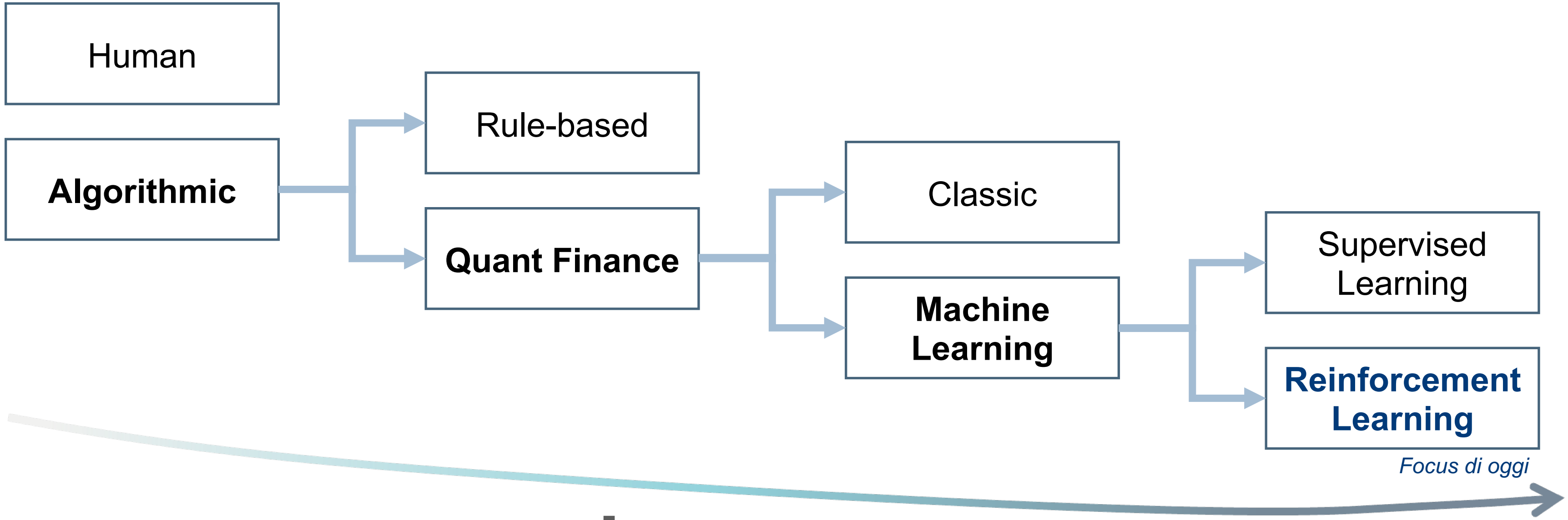
Il mercato del **trading algoritmico** cresce con un **CAGR** del **~11%** ('21-'26)

Principali tipologie di algoritmi

- **Optimal execution** e smart routing
- Market making
- Hedging
- **Trading**
- Portfolio optimization

Tecnologie di Trading Algoritmico

Classificazione per tipologia di tecnologia



+ Indipendenza dall'uomo

+ Complessità computazionale

+ Performance

Algorithmic Trading

Introduzione

Reinforcement Learning

- **Tecnica di apprendimento automatico** per realizzare **algoritmi in grado di scegliere** le **azioni** in maniera autonoma
- Algoritmo impara a raggiungere un **obiettivo** tramite **interazione** con l'**environment**

Breakthrough del Reinforcement Learning

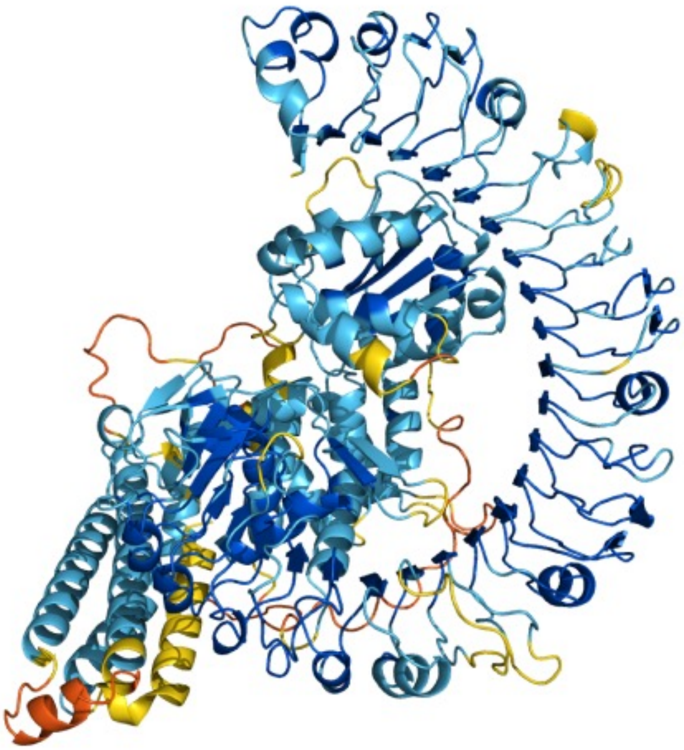
AlphaGo

2016

AlphaGo batte il campione di Go

Breakthrough del Reinforcement Learning

AlphaFold



2020

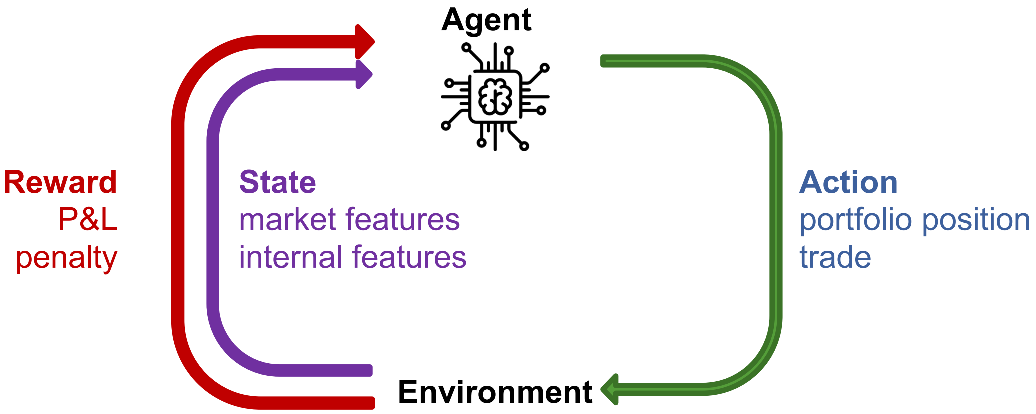
AlphaFold risolve una sfida aperta da 50 anni nella modellazione delle proteine

Ulteriori Applicazioni del Reinforcement Learning

Guida autonoma e robotica

Basi del Reinforcement Learning

Markov Decision Process: processo per descrivere interazione tra agente ed environment



- Obiettivo è trovare la policy π che massimizza la somma scontata dei reward
- $J = \max_{\pi} \mathbb{E}_t[\sum \gamma_t R_t]$

Reinforcement Learning per il Trading

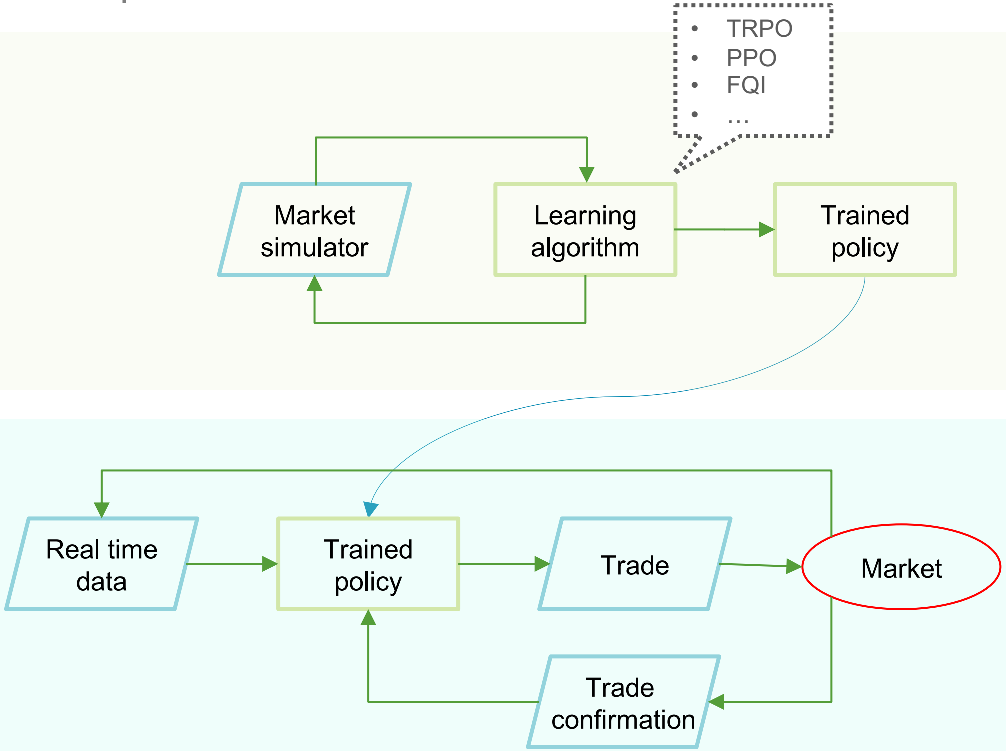
Training, testing ed utilizzo in produzione

Fase 1

- Training
- Hyperparameter tuning
- Backtesting

Fase 2

- Produzione



Algorithmic Trading

Reinforcement Learning per Quantitative Trading

Definizione problema e descrizione MDP

Quantitative Trading

Definizione

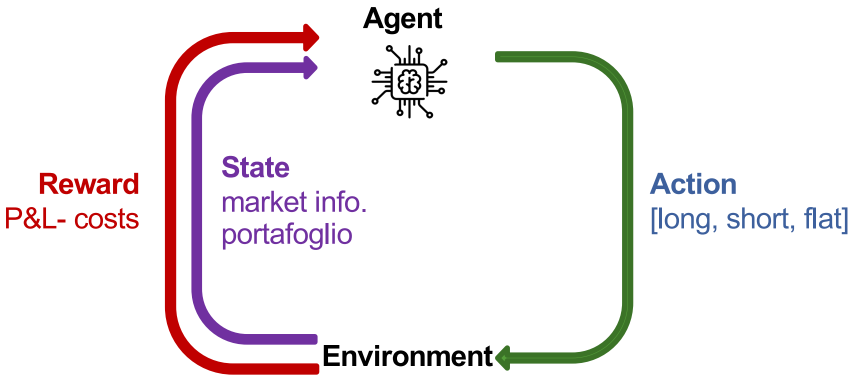
- Ad ogni time-step, decidere se andare long, short o flat per massimizzare il guadagno

MDP

- **State:** finestra di prezzi, bid-ask spread, portafoglio corrente, data/ora
- **Action:** long, short, flat
- **Reward:** P&L – costi di transazione

Caratteristiche

- Alpha seeking
- Bassa correlazione con il mercato



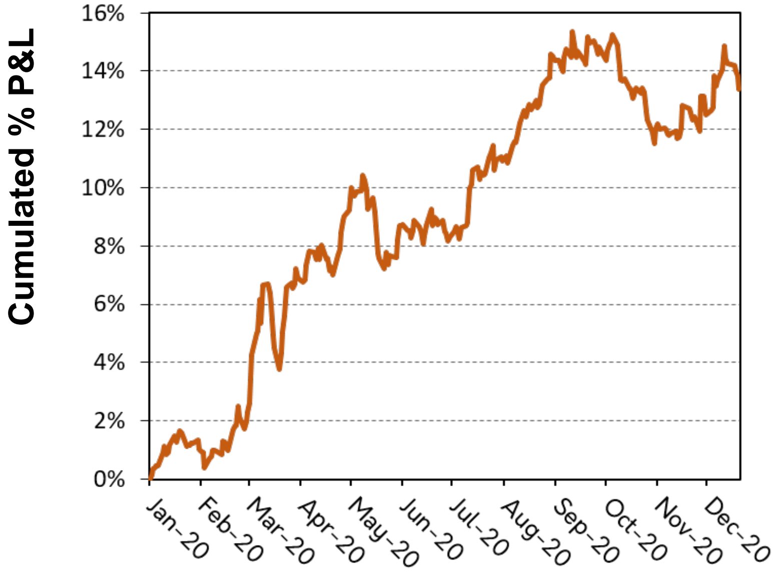
Reinforcement Learning per FX Trading (1/2)

Risultati sperimentali - performance

Esperimento

- Intraday trading on EURUSD FX
- Training con FQI su dati storici 2017-2018
- Validation su dati storici 2019
- Backtesting su dati storici out of sample 2020

P&L of Backtest EURUSD FX trading su 2020



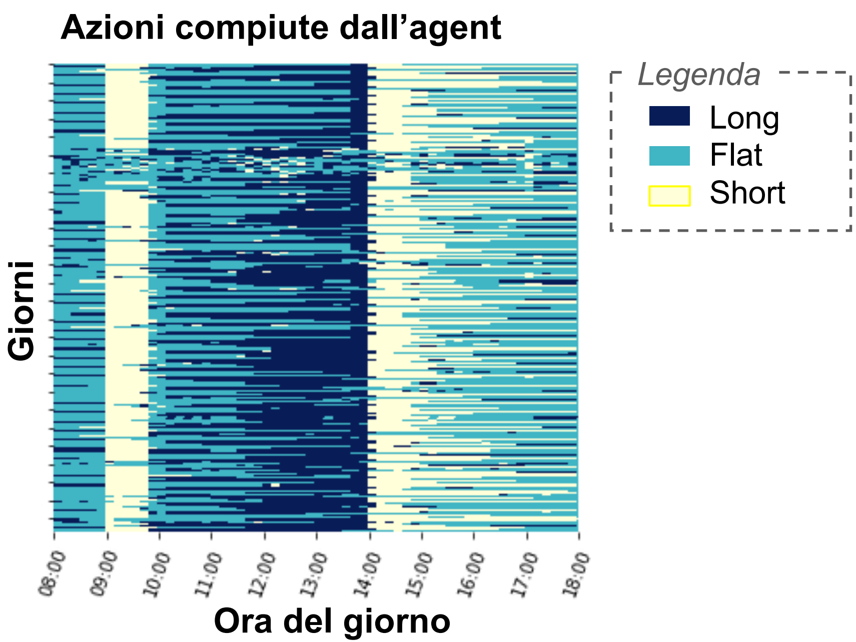
Riva A., Bisi L., Liotet P., Sabbioni L., Vittori E., Trapletti M., Pinciroli M., Restelli M., 2021

Reinforcement Learning per FX Trading (2/2)

Risultati sperimentali - policy

Esperimento

- Intraday trading on EURUSD FX
- Training con FQI su dati storici 2017-2018
- Validation su dati storici 2019
- Backtesting su dati storici out of sample 2020



Riva A., Bisi L., Liotet P., Sabbioni L., Vittori E., Trapletti M., Pinciroli M., Restelli M., 2021

Se consideriamo impatti sul mercato?

- Finora abbiamo assunto di avere **costi di transazione ma nessun impatto** sul mercato
- Che succede se abbiamo **impatti**?

?

Algorithmic Trading

Limit Order Book

Definizione ed esempio di limit order book

Caratteristiche

- Limit order book è il record di tutti i limit order non eseguiti
- Limit order è un ordine che specifica prezzo e quantità del trade
- Market order è un ordine di eseguire il trade immediatamente al migliore prezzo possibile

Esempio Limit Order Book

Last	Last Vol	Total Vol	Close	Daily Low	Daily High
4045.00	2	367267	4097.50	4033.50	4101.50

Implied

Bid		Offer	
Volume	Price	Price	Volume
136	4044.50	4045.00	62
327	4044.00	4045.50	293
348	4043.50	4046.00	427
620	4043.00	4046.50	426
358	4042.50	4047.00	463
330	4042.00	4047.50	348
325	4041.50	4048.00	327
318	4041.00	4048.50	294
305	4040.50	4049.00	281
512	4040.00	4049.50	288

Reinforcement Learning per Optimal Execution

Definizione problema e descrizione MDP

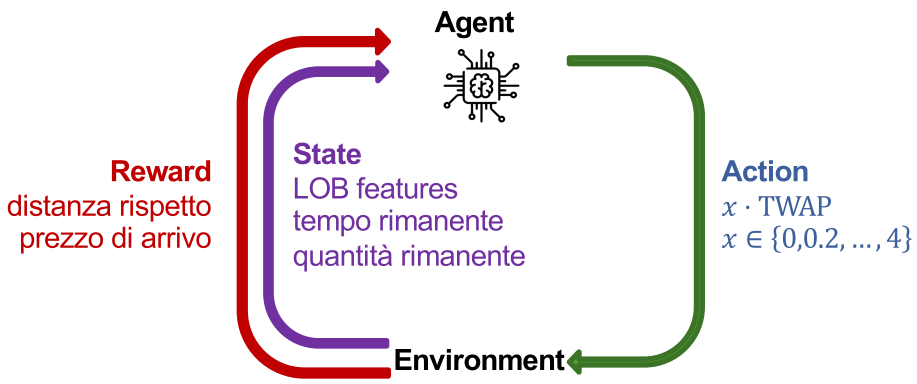
Optimal Execution

Definizione

- Eseguire X shares in N timesteps
- Decidere ad ogni timestep il trade da eseguire per minimizzare differenza tra prezzo di arrivo e di esecuzione

MDP

- **State:** features del LOB, timestep rimanenti, quantità rimanente
- **Action:** x·TWAP con $x \in \{0, 0.2, \dots, 4\}$
- **Reward:** distanza rispetto prezzo di arrivo



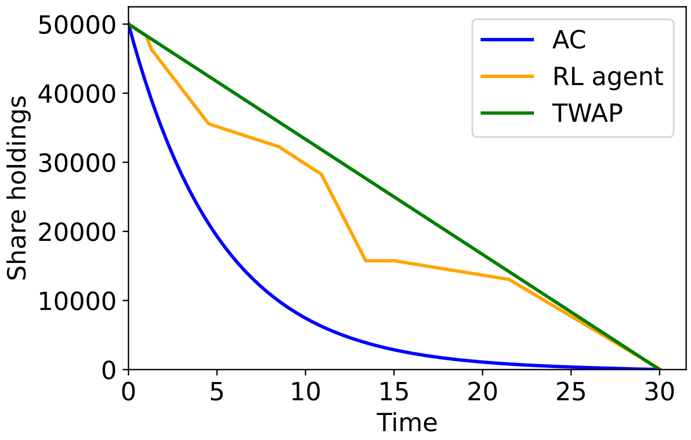
Risultati Sperimentali

Paragone di return tra RL e benchmark su mercato simulato con ABIDES

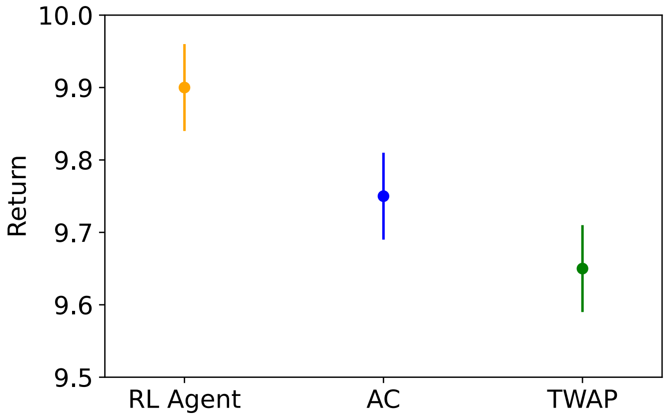
Caratteristiche

- Simulazione con ABIDES di un esperimento di optimal execution (ripetuto più volte)
- 30 minuti per eseguire 50k shares

Traiettorie di esecuzione



Return medio agente RL vs benchmarks



Conclusioni

Real Estate

Utilizzi ML

- Stima valore immobili
- Predizione trend di mercato

Benefici

- Riduzione sensibile dei costi
- Riduzione dei tempi di analisi

Credit Risk Management

Utilizzi ML

- Feature selection
- Data Clustering
- Performance enhancing

Benefici

- Integrazione di dati micro e macro economici
- Stima rischio di credito da fatture e transazioni

Algorithmic Trading

Utilizzi ML

- Quantitative trading
- Optimal Execution

Benefici

- Creazione strategie profittevoli a bassa correlazione con il mercato
- Adattamento dinamico alle condizioni del mercato

Machine Learning for Finance

Applicazioni in Credit Scoring, Real Estate e Trading Algoritmico

Emilio Colombo
emilio.colombo@cerved.com

Cristian Giacomini
cristian.giacomini@modefinance.com

Edoardo Vittori
edoardo.vittori@master.sdabocconi.it