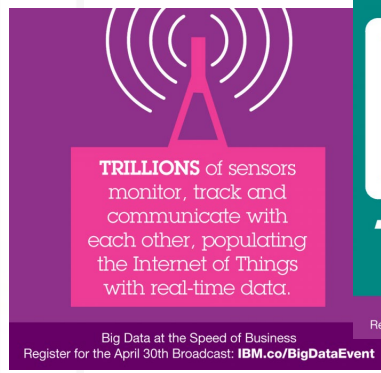
AllianceBernstein

Leveraging Graph Theory and NLP to Perform Causal Inference from News Articles

May 2024

Andrew Chin – Head of Investment Solutions and Sciences
Che Guan – Principal Data Scientist

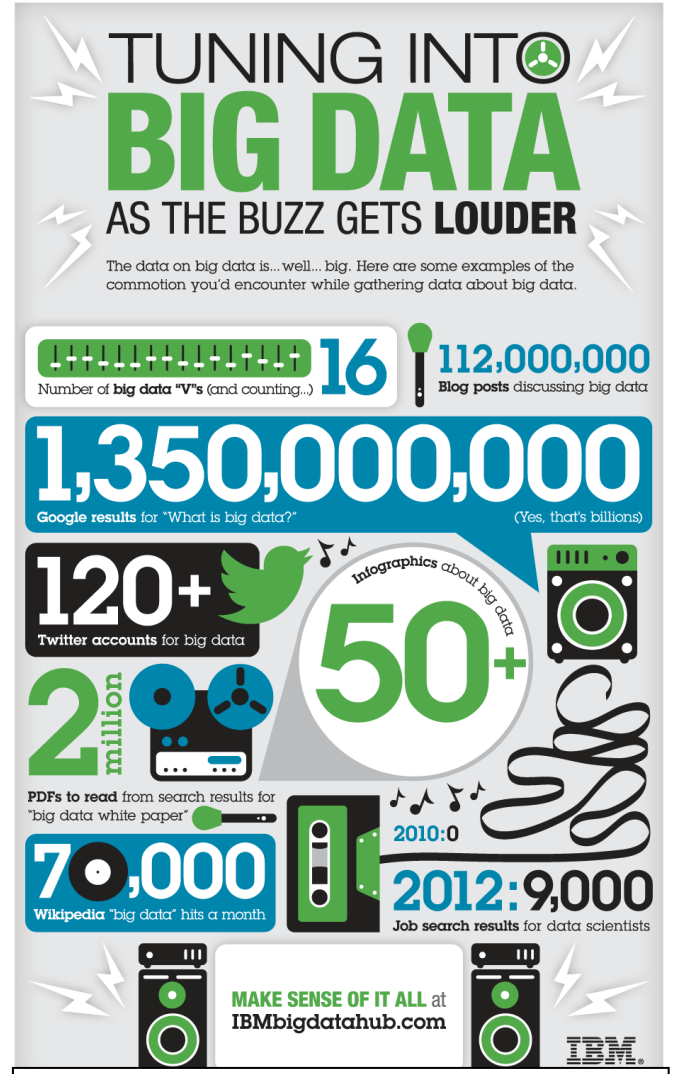
A Lot More Data Available!

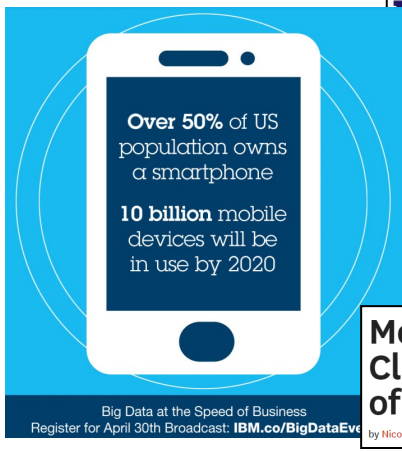


How Big Data Is Used To Fight Cyber Crime And Hackers: Fascinating Use Case From BT

Bernard Marr, CONTRIBUTOR

Forbes





InformationWeek | Join us live at Interop ITX

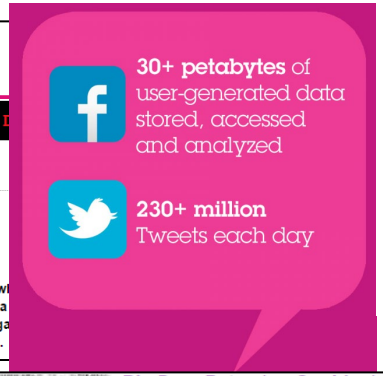
IT Leadership | DevOps | Security | Cloud

DATA MANAGEMENT // BIG DATA ANALYTICS

NEWS 12/2016 9:06 AM

Big Data Moves From Hype To Reality, CompTIA Finds

Organizations are moving past the hype and into actual value wh comes to big data and analytics implementation, according to a survey by CompTIA. But challenges remain, including a skills ga rated.



Big Data = Big Success At Top Investment Fund

Forbes

Most Industries Are Nowhere Close to Realizing the Potential of Analytics

by Nicolaus Henke, Jacques Bughin, and Michael Chui

Harvard Business Review

Big Data Poised to Get Much Bigger in 2017

Gigaom - Dec 1, 2016

Big Data is only going to get much bigger, so big in fact that companies ... insights and **statistics** on how **Big Data** is poised to change in 2017. ... at its October Business Intelligence & Analytics Summit **2016**, in Munich, that ...

Artificial Intelligence / Machine Learning to the Rescue?

Forbes

MONEY

How AI And ML Are Changing Finance In 2022

Dmitry Dolgorukov Forbes Councils Member
Forbes Finance Council
COUNCIL POST | Membership (Fee-Based)

Dec 17, 2021, 07:00am EST

Dmitry Dolgorukov is the Co-Founder and CRO of HES Fintech, a leader in providing financial institutions with intelligent lending platforms.

GETTY

2021 was a year marked by the implementation of the rapid digital transformations that first sprouted when the coronavirus pandemic hit the world in 2020. Fintech firms and other businesses around the world invested heavily in transforming to meet the needs of the new normal — remote working, social distancing and a business world changed perhaps forever.

FORTUNE

Great Resignation | Climate Change | Leadership | Inflation | Ukraine Invasion

COMMENTARY · CLIMATE CHANGE

The power of A.I. to help mitigate and manage climate change

BY CHRISTOPH SCHWEIZER
September 19, 2022 at 3:00 PM EDT

Finance Monthly

Finance News | Finance Monthly Interview | Most Recent | Opinion & Analysis | Financia

Applications Of Machine Learning In Banking Risk Management

Artificial intelligence has been finally recognised as the technology that can transform banks' critical functions.

From chatbots to credit underwriting to stock market predictions, there is no shortage of use cases of machine learning in banking.

Despite the fact that risk management has always been at the top of banks' agenda, many processes are still plagued with inefficiencies that are continuously draining bank resources. In this article, Andrey Koptelov discusses how banks can apply machine learning to streamline regulatory risk management and advance their fraud detection methods.

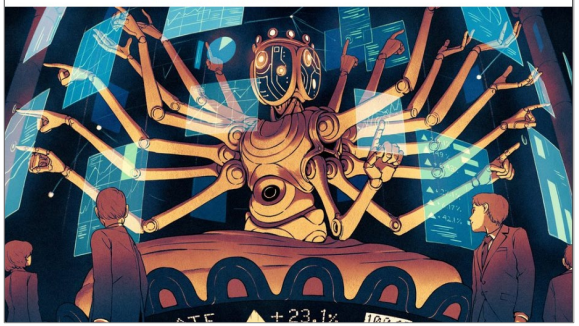
Bloomberg

Markets

How AI Will Invade Every Corner of Wall Street

Machine learning, with its prowess in producing insights from data, is poised to have a hand in 99 percent of investing, CEO says.

By Nishant Kumar
December 5, 2017 at 2:00 AM EST



Nasdaq

MARKET ACTIVITY | NEWS + INSIGHTS | SOLUTIONS

TECHNOLOGY

How Big Data and AI Are Changing the Financial Industry

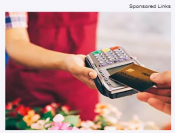
CONTRIBUTOR David Cotriss

PUBLISHED JUN 23, 2022 9:02AM EDT

CREDIT: SHUTTERSTOCK

The financial industry has never been short of data, but until recently, the bulk of it has been too complex to do anything meaningful with. A little something called AI is gradually changing that. But what impact will that have on the financial industry, and which companies should we be watching?

Sponsored Links

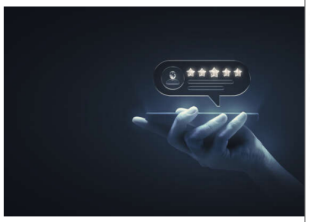


6 Credit Cards You Should Not Ignore If You Have Excellent Credit

Finance

TECHNOLOGY

How AI and ML testing and other techniques help maintain finance quality



By Eran Kinsbruner, DevOps Chief Evangelist, Perforce Software

Software development has come a long way in recent years, with DevOps, continuous delivery/continuous integration pipelines, increased automation and more 'X-as-code' helping to keep up with the demand to create ever larger volumes of increasingly complex code. However, the need to test it thoroughly is just as critical as creating that code. Otherwise, the velocity gains are negated by the risk of flaws, which can cause performance issues or even be a backdoor for cyberattacks and data breaches. Clearly, in the financial services world, secure software has to be a priority for regulatory requirements and customer satisfaction.

TC

Join TechCrunch+
Login

Search
TechCrunch+
Startups
Venture
Security
Crypto
Apps
Events
Advertise
More

Arthur.ai machine learning monitoring gathers steam with $42M investment

Ron Miller @ron_miller / 10:00 AM EDT • September 27, 2022

Comment

Image Credits: Oscar Wong / Getty Images

It's widely understood that after machine learning models are deployed in production, the accuracy of the results can deteriorate over time. Arthur.ai launched in 2019 with the goal of helping companies monitor their models to ensure they stayed true to their goals. Since then, the company has also added explainability and bias mitigation to the array of services.

CISION

PR Newswire

Send a Release

Machine Learning in Banking Market to Garner $21.27 Billion, Globally, By 2031 at 32.2% CAGR: Allied Market Research

Allied Market Research

NEWS PROVIDED BY
Allied Market Research →
Sep 26, 2022, 11:30 ET

SHARE THIS ARTICLE

Improved productivity of banks and faster banking operations using machine learning have boosted the growth of the global machine learning in banking market.

PORTLAND, Ore., Sept. 26, 2022 /PRNewswire/ -- Allied Market Research recently published a report, titled, "**Machine Learning in Banking Market By Component (Solution and Service), Enterprise Size (Large Enterprises, Small and Medium-Sized Enterprises [SMEs]), Application (Credit Scoring, Risk Management Compliance and Security, Payments and Transactions, Customer Service, and Others): Global Opportunity Analysis and Industry Forecast, 2021-2031**". As per the report, the global machine learning in banking industry accounted for $1.33 billion in 2021, and is expected to reach $21.27 billion by 2031, growing at a CAGR of 32.2% from 2021 to 2030.

Extracting Insights from Unstructured Data

- News and reviews
- Twitter / social media
- Sentiment analysis
- Corporate filings
- Central Bank statements
- Web searches/activity
- App usage
- User/client data
- Patents

- Web-scraping (prices, inventories, activities)
- Transaction data
- Email receipts
- Web clicks (for client interactions)
- Business data (job listings, employee reviews)
- Government data

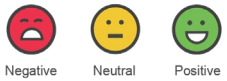
- Location data (stores, suppliers, customers, etc)
- Shipping/logistics data
- Satellite images
- ESG-related data
- Covid-related

- Health/insurance data
- Internet of Things
- Proprietary data from internal investors, sales force and operational groups
- Synthetic data from LLMs

Common AI Tasks in Asset Management

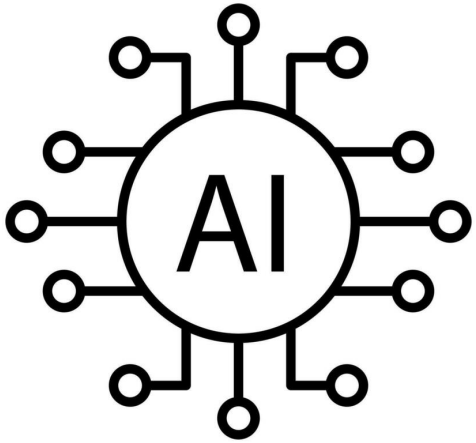
Interpretation

Named Entity Recognition

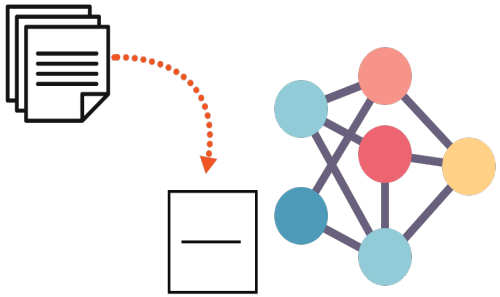


Sentiment Analysis

Prediction

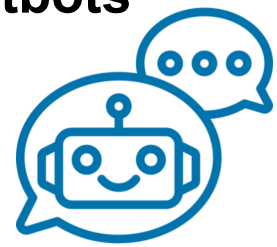


Summarization / Topic Modeling



Search / Question Answering

Chatbots



Content Creation

Our Project – Predict Impact of News Articles on Stock Prices

- Identify causal events in news articles
- Construct graph to capture impacts of events
- Use graph to infer relationships on test dataset to assess effectiveness

Task #1 - Annotate Event, Event-Company Relationship, and Impact in News Articles

- Leverage LLM to identify and highlight key events mentioned in each article, giving particular emphasis to the following categories: defaults, mergers and acquisitions, revenue discussions, margin/profitability discussions, industry competitiveness
- Create three labels for each event:
  - Identify and distinguish between causal and non-causal events within the article, as well as potentially across multiple articles.
  - Identify and capture the relevant companies/assets mentioned or impacted by the events. Multiple identities per article, such as stocks, bonds, commodities, etc., are permissible.
  - Assess and quantify the impact of each event on the labeled company's stock performance and associated asset class(es), categorizing it as positive, negative, or neutral.
- Create an annotation dataset from at least 5,000 news articles for the three labels mentioned above, it is necessary to carefully evaluate the accuracy for the initial 500 annotations. If the accuracy falls below 85%, it is recommended to adjust the prompt, e.g., employing few-shot learning or fine-tuning the LLM to enhance accuracy. However, if the accuracy is satisfactory, we can proceed confidently with the labels provided by the LLM, while still incorporating human evaluation and correction.

Task #2 - Graph-based Representation and Analysis of Graph Features

- Construct a directed graph, where the nodes represent news events and companies/assets, and the directed edges represent causal links. The direction of the edges indicates the flow or order of causality between the nodes.
- Assign weights to the edges based on the confidence level of the causal relationships. Stronger connections, indicating more significant impacts, will be assigned higher weights.
- Analyze and visualize various graph properties, such as the importance of nodes and edges, as well as identify self-contained subgraphs that consist of causally linked events and associated companies/assets.

Task #3 - Linkage Prediction

- Utilize graph neural networks, such as Graph Convolutional Networks (GCN), GraphSAGE, and Graph Attention Networks (GAT), to generate embeddings for each node in the graph. This involves iteratively learning embeddings from the node's neighbors and itself, capturing the underlying relationships. Leverage the learned embeddings to infer the linkage between news events and the identified companies/assets in the graph.
- Apply either inductive link prediction split or transductive link prediction split techniques to enhance the accuracy of link prediction within the graph.
- Explore the use of Knowledge Graph (KG) and KG completion techniques to further enrich the graph representation and improve the understanding of causal relationships within data.

Task #4 - Derive and Backtest Investment Signal(s)

- Perform analysis on a selected set of events across a wide range of companies, prioritizing breadth over depth. The focus will be on examining a few events that have occurred across multiple companies.
- Derive investment signals by leveraging the impact scores assigned to each event and each company per day. Utilize this information to construct a long-short portfolio.
- Evaluate the performance of the portfolio by considering open-close prices over different time horizons and sectors, such as 1 day, 5 days, and 10 days.

Expectations and Deliverables

- **Project Deliverables**
  - ~5K annotations
  - Well-documented models/pipeline which can be directly used by AllianceBernstein
  - A technical report describing project specifics, e.g., documenting data pre-processing steps, LLM prompts, various graph models, details of experiments conducted, necessary steps for reproducing projects, etc.
  - Final presentation to investment teams in Dec 2024
- **Success Metrics**
  - Sentiment Classification: Precision, Recall, and F1 Score
  - Signal Alpha: Excess return against the benchmark (equally weighted S&P 500 return),
- **Project Logistics**
  - Jun – Jul 2024: Tasks 1 and 2
  - Aug – Sep 2024: Tasks 3 and 4
  - Weekly project catchups:
    - Team leader with agenda for each meeting
    - List of take-aways and follow-ups for next meeting

References

- Andrew Chin and Yuyu Fan “Leveraging Text Mining to Extract Insights from Earnings Call Transcripts,” Journal Of Investment Management, Vol. 21, No. 1, (2023), pp. 81–102
- Price, Chris. "Estimating Causal Effects on Financial Time-Series with Causal Impact BSTS." Towards Data Science. January 29, 2020
- CS224W | Home (stanford.edu)
- https://medium.com/@adrian.mladenic.grobelnik/predicting-impactful-events-with-graphml-ccd0feefa869
- https://www.youtube.com/watch?v=1RZ5yIyz31c