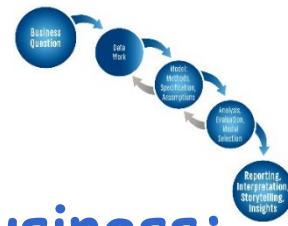


Leadership Initiatives AI & Emerging Technology Internship



Machine Learning & Artificial Intelligence for Business: Key Concepts and Trends

Prof. J. Alberto Espinosa, PhD

Dept. of Information Technology & Analytics

alberto@american.edu

July 28, 2025



Slides: <https://tinyurl.com/EspinosaLead>



[Amazon](#)



[Ch1](#)



[R: GitHub](#)



[ChatBot](#)

PAML4M Book



KOGOD SCHOOL *of* BUSINESS
AMERICAN UNIVERSITY • WASHINGTON, DC

J. Alberto Espinosa, Ph.D.
Professor of Information Technology & Analytics
Kogod School of Business, American University, Washington, D.C.

Prof. Espinosa holds a *Ph.D.* and *Master of Science in Information Systems* from the *Tepper School of Business* at *Carnegie Mellon University*, a *Master of Business Administration* from *Texas Tech University*, and a *Mechanical Engineering* degree from *Pontificia Universidad Católica del Perú*. He is the architect and implementor of *Kogod's MS Analytics Program* and the undergraduate specialization in *Business Analytics*. He is also the curriculum architect for the *IT&A* programs and specializations. He has over 70 published journal articles, conference proceedings, books, and book chapters. He has co-authored two research books, one on work coordination across time zones and a 2-volume book on big data and analytics for service delivery. He has also written a book titled "*Predictive Analytics and Machine Learning for Managers*" with practical applications in R. His research focuses on coordination and performance in technical projects across global boundaries, mainly distance and time zones. More recently, he has developed social network and graph analytics methods to represent team knowledge quantitatively and visually. His work has been published in leading scholarly journals, including *Management Science*, *Organization Science*, *Information Systems Research*, the *Journal of Management Information Systems*, *IEEE Transactions on Software Engineering*, *IEEE Transactions on Engineering Management*, *Communications of the ACM*, *Human Factors, Information, Technology and People*, and *Software Process: Improvement and Practice*. His work has also been featured in leading academic conference proceedings. He teaches predictive analytics with machine learning, social & organizational network analytics, R programming, information technology foundations, business process analysis, and programming for business applications. He has received multiple service, research, and teaching awards, including "American University 2025 Outstanding Teacher", "2022 Kogod Graduate Professor of the Year", and "American University 2022 Green Teacher". He has several years of work experience, first as a mechanical design engineer and later as a senior manager, VP, and CFO with international organizations directly supporting, supervising, and formulating policy for finance and global IT, data management and analysis, where he designed and developed many software applications to support geographically distributed work.

A Story about Estor - AI is Not New

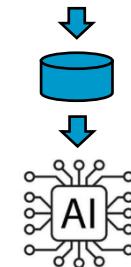
💡 **What was Estor?** It is one of the earliest expert systems applied to software effort estimation, developed by Steven Vicinanza, a Carnegie Mellon PhD student in 1990, and published in MIS Quarterly in June of 1992, “*Examining the Feasibility of a Case-Based Reasoning Model for Software Effort Estimation*”

- It was built by analyzing how **experienced** software project estimators reasoned when estimating time to completion (i.e., effort) in real software projects.



⚙️ **How did it work?** When estimating new projects, these experts retrieved similar past cases and applied adaptation **if-then rules**. These rules were coded into a **rule database** to be queried by the **Expert System**, mirroring the expert's thought process.

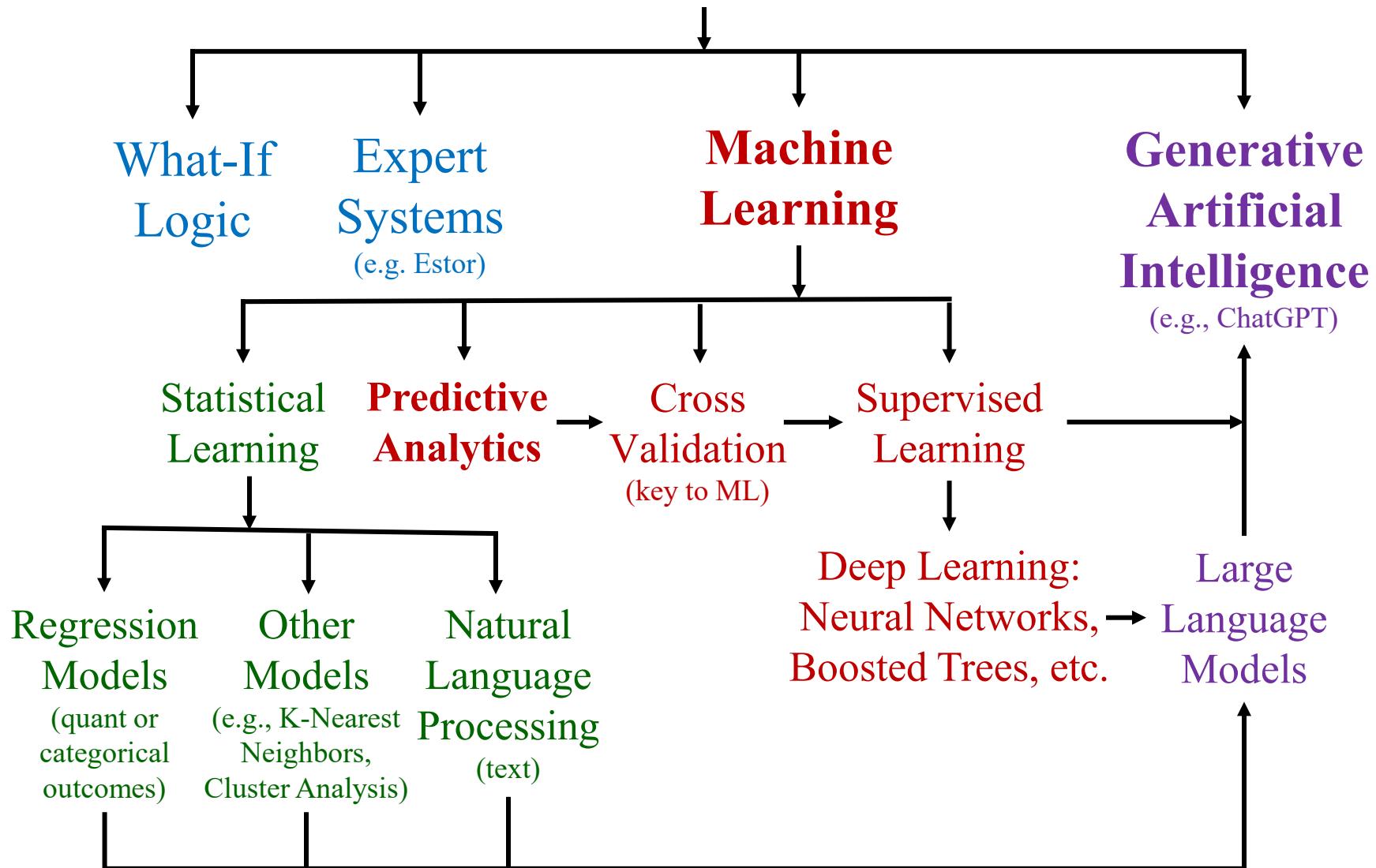
📊 **Results:** Estor was tested on 15 real software projects, comparing its estimates against those of the experts and various quantitative estimation methods. **Estor** was as accurate as the human experts and more accurate than the traditional estimation methods.



✳️ **Importance:** Demonstrated the feasibility of expert system AI for software estimation back in the **early 1990s**, when machine learning was not mainstream, which inspired over three decades of follow-up research using machine learning, neural networks, regression, and fuzzy systems to improve estimation accuracy.

In summary, **Estor** was a groundbreaking early **AI effort estimator** using insights from expert estimators. It proved that capturing expert thinking in a structured rule base could match human-level estimates, setting the stage for decades of research in this field.

Main Branches of AI



Predictive Analytics

THE WALL STREET JOURNAL.

WSJ.com

September 11, 2015, 9:28 AM ET

Facebook ‘Likes’ Mean a Computer Knows You Better Than Your Mother

By Georgia Wells



Multiple studies over the years have confirmed that “likes” can predict traits, preferences, personality, etc., with 85% to 95% accuracy.

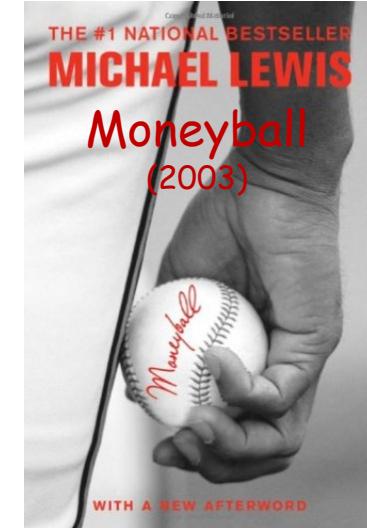
A computer needs the data from only 10 Facebook “likes” to beat the accuracy of a person’s coworker at judging his or her personality traits, such as extraversion, conscientiousness and neuroticism. The computer needs 70 “likes” to be more accurate than a person’s friends. With 250 Facebook data points, the computer can beat someone’s spouse.



Business Analytics

“It is the scientific process of transforming data into insights for making better decisions” (INFORMS)

Lack analytical talent – by 2018, the US would face a shortage of **140,000** to **190,000** people with **deep analytical skills** and **1.5 million managers** with analytical skills to make decisions with data (McKinsey Global Institute, 2011)



The Age of Analytics

McKinsey Global Institute

2016

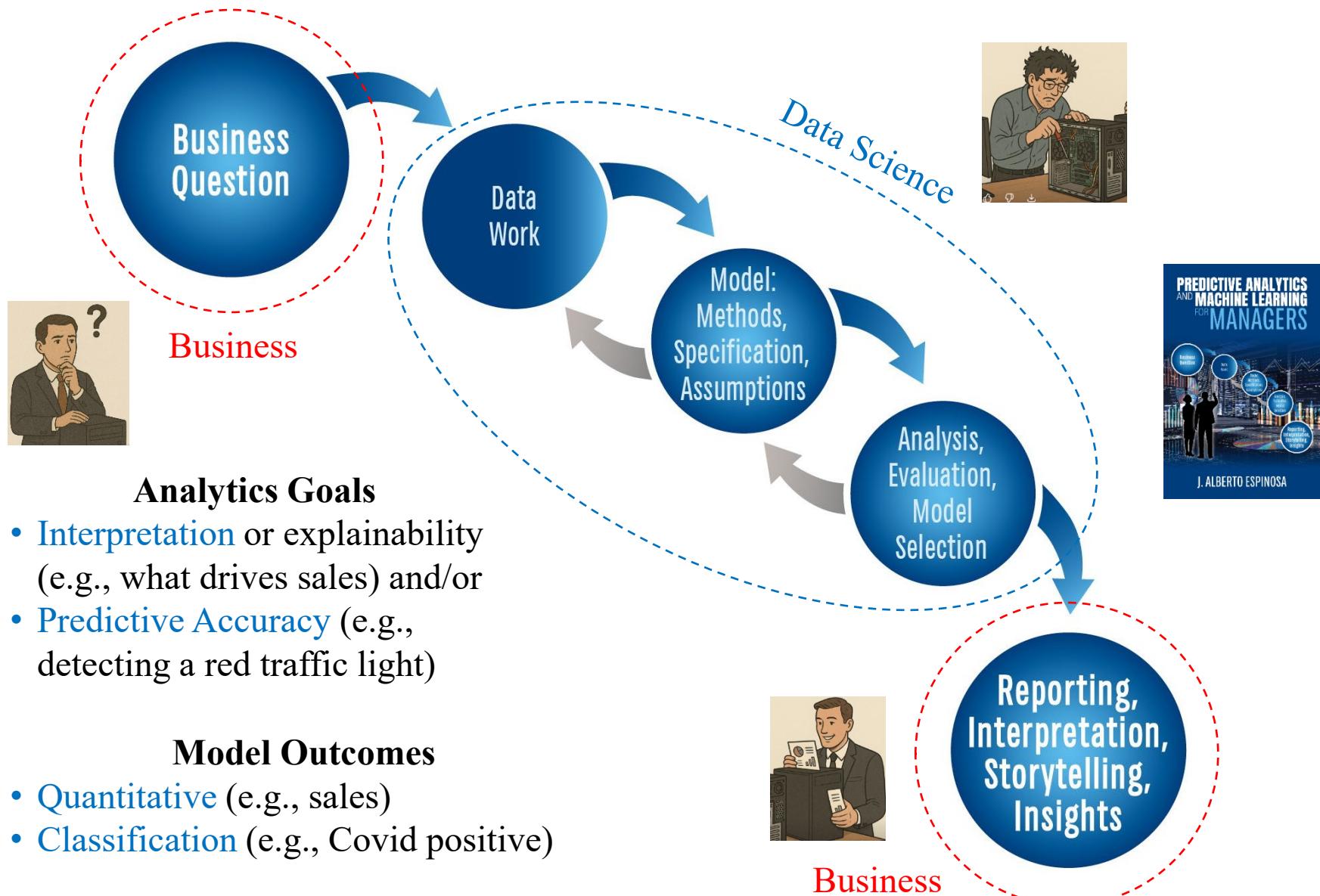
2M-4M

Projected US demand for “business translators” in the **2020** decade

High demand for **business translators** → people who serve as the link between **data science** talent and business applications and **managers**, who are business-savvy and can answer business questions.

In 2025 McKinsey issued their “AI in the Workplace” report, stating that employers are eager to adopt AI, but leadership and readiness in this area lag behind → Almost **all** companies invest in **AI**, but just **1%** believe they lack **leaders** who understand AI.

The Analytics Process



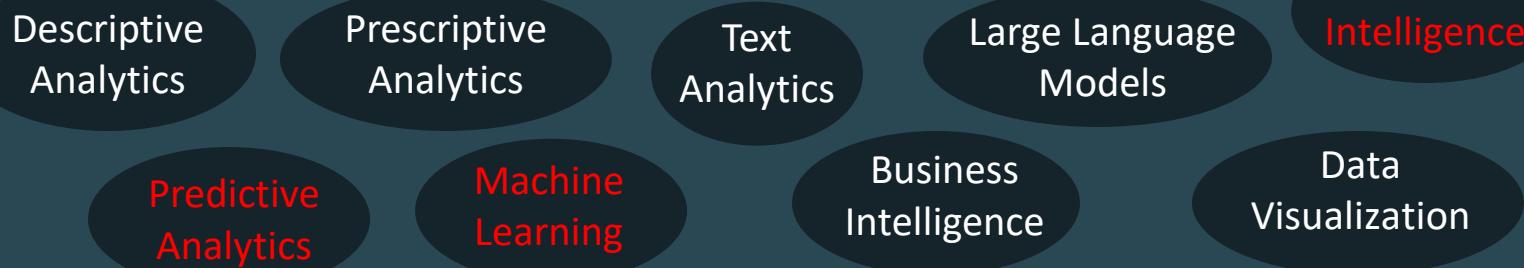
Analytics Body of Knowledge



Business Domains



Analytics & Data Science



Foundations (IT & Math)

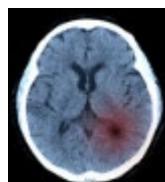
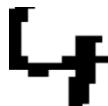


Machine Learning

Is method of data analysis that uses algorithms that **learn from data** iteratively, allowing computers to find hidden insights **without being explicitly programmed**

(SAS: http://www.sas.com/en_us/insights/analytics/machine-learning.html)

- **Unsupervised learning** → data exploration **without** specific **goals** (closely associated with descriptive analytics and data mining, e.g., clustering, correlation analysis, histograms)
- **Supervised learning** → data analysis **with** specific **goals** in mind (closely associated with analytics, e.g., regression) → not much different than **predictive modeling**



\$600

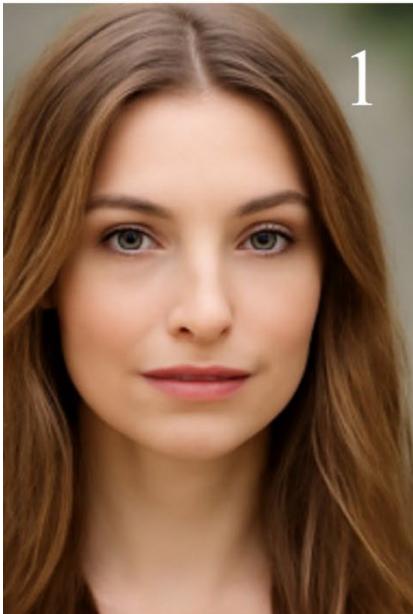


S200 OFF
Miele CM7750 Fully
Automatic Espresso Machine

\$5,300



Classify each person as “hot” or not to you:



LooksMapping app by a 22-year old developer to rate restaurants on a “hotness” scale

Predictive Model

With Machine Learning, without Cross Validation

| <i>Predictors</i> | | | | | | | | \rightarrow | <i>Outcome</i> |
|-------------------|-------|-------|-------|-------|------|--------|-------|---------------|----------------|
| PizzaID | brand | mois | prot | fat | ash | sodium | carb | | |
| 14001 | D | 47.17 | 22.29 | 21.3 | 4.08 | 0.74 | 5.16 | | |
| 14002 | D | 49.16 | 27.99 | 17.49 | 3.29 | 0.39 | 2.07 | | |
| 14003 | A | 30.49 | 21.28 | 41.65 | 4.82 | 1.64 | 1.76 | | |
| 14004 | B | 52.68 | 14.38 | 25.72 | 3.26 | 0.93 | 3.96 | | |
| 14005 | H | 33.05 | 7.34 | 15.78 | 1.34 | 0.42 | 42.49 | | |
| 14006 | H | 35.55 | 7.32 | 16.4 | 1.76 | 0.36 | 38.97 | | |
| 14007 | G | 28.68 | 8.3 | 16.07 | 1.41 | 0.45 | 45.54 | | |
| | | | | | | | | cal | |
| | | | | | | | | 302 | |
| | | | | | | | | 278 | |
| | | | | | | | | 467 | |
| | | | | | | | | 305 | |
| | | | | | | | | 341 | |
| | | | | | | | | 333 | |
| | | | | | | | | 360 | |

Machine Learning w/Cross-Validation

→ Split data set + Train model + Test it

Training Subset: *Train Model* → *Outcome*

| PizzaID | brand | mois | prot | fat | ash | sodium | carb |
|---------|-------|-------|-------|-------|------|--------|------|
| 14001 | D | 47.17 | 22.29 | 21.3 | 4.08 | 0.74 | 5.16 |
| 14002 | D | 49.16 | 27.99 | 17.49 | 3.29 | 0.39 | 2.07 |
| 14003 | A | 30.49 | 21.28 | 41.65 | 4.82 | 1.64 | 1.76 |
| 14004 | B | 52.68 | 14.38 | 25.72 | 3.26 | 0.93 | 3.96 |

| |
|-----|
| cal |
| 302 |
| 278 |
| 467 |
| 305 |

Test Subset: *Test Model's Accuracy* →

| | | | | | | | |
|-------|---|-------|------|-------|------|------|-------|
| 14005 | H | 33.05 | 7.34 | 15.78 | 1.34 | 0.42 | 42.49 |
| 14006 | H | 35.55 | 7.32 | 16.4 | 1.76 | 0.36 | 38.97 |
| 14007 | G | 28.68 | 8.3 | 16.07 | 1.41 | 0.45 | 45.54 |

| | | |
|-----|-----|-----|
| 341 | 290 | 51 |
| 333 | 362 | -29 |
| 360 | 371 | -11 |

Actual Predicted Error

Mean Squared Error: 1187.7

RMSE = 34.5

Selected Model

| | | | | | | | |
|----------|---|-------|-------|-------|------|------|------|
| New Data | D | 49.16 | 27.99 | 17.49 | 3.29 | 0.39 | 2.07 |
|----------|---|-------|-------|-------|------|------|------|

Uses: test and compare models, tune model parameters, train deep and supervise learning models, etc.



Predict:

- Values
- Classifications
- Words
- Images, etc.

Machine Learning Models (e.g., Neural Networks)

- Simple learning → simple linear relationships & 1 or 2 layers
- Deep learning → linear or non-linear relationships & multiple layers

ex. recognize
hand-written
zip code
numbers

Input Layer



ex., recognize
a traffic light &
decide to stop
or not

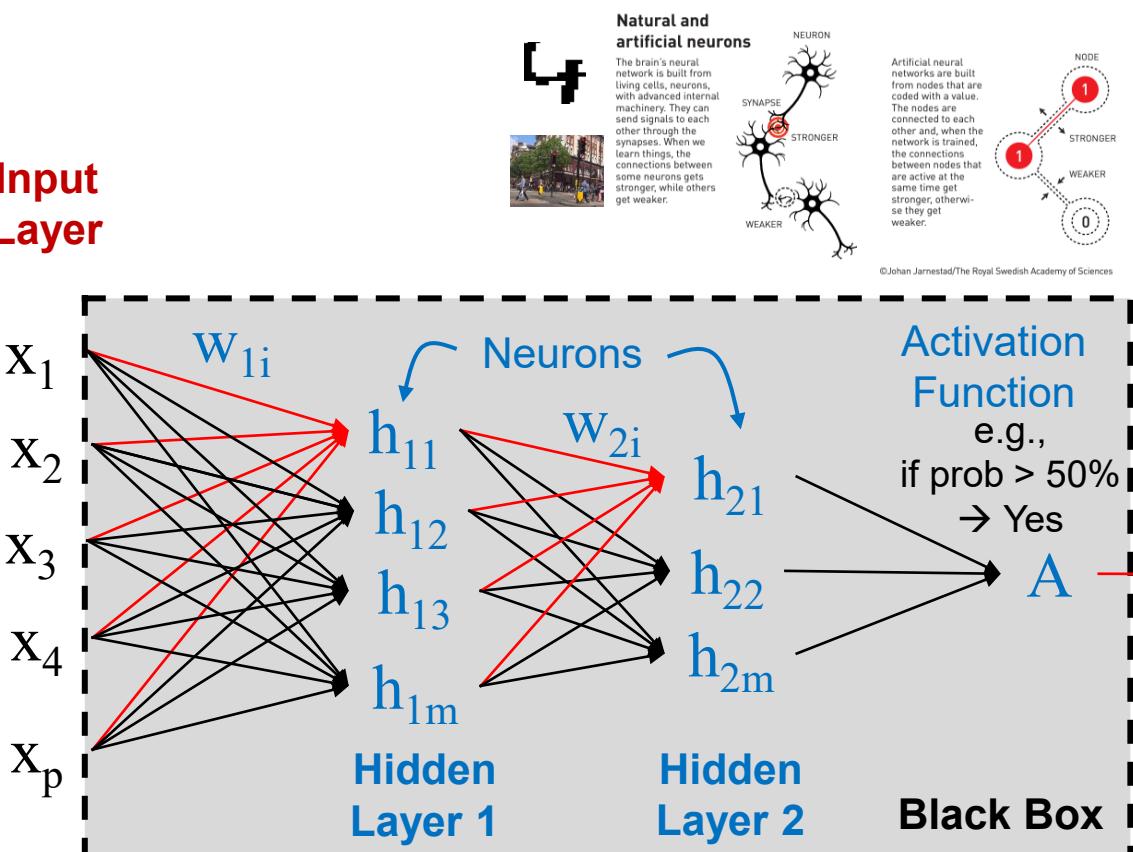
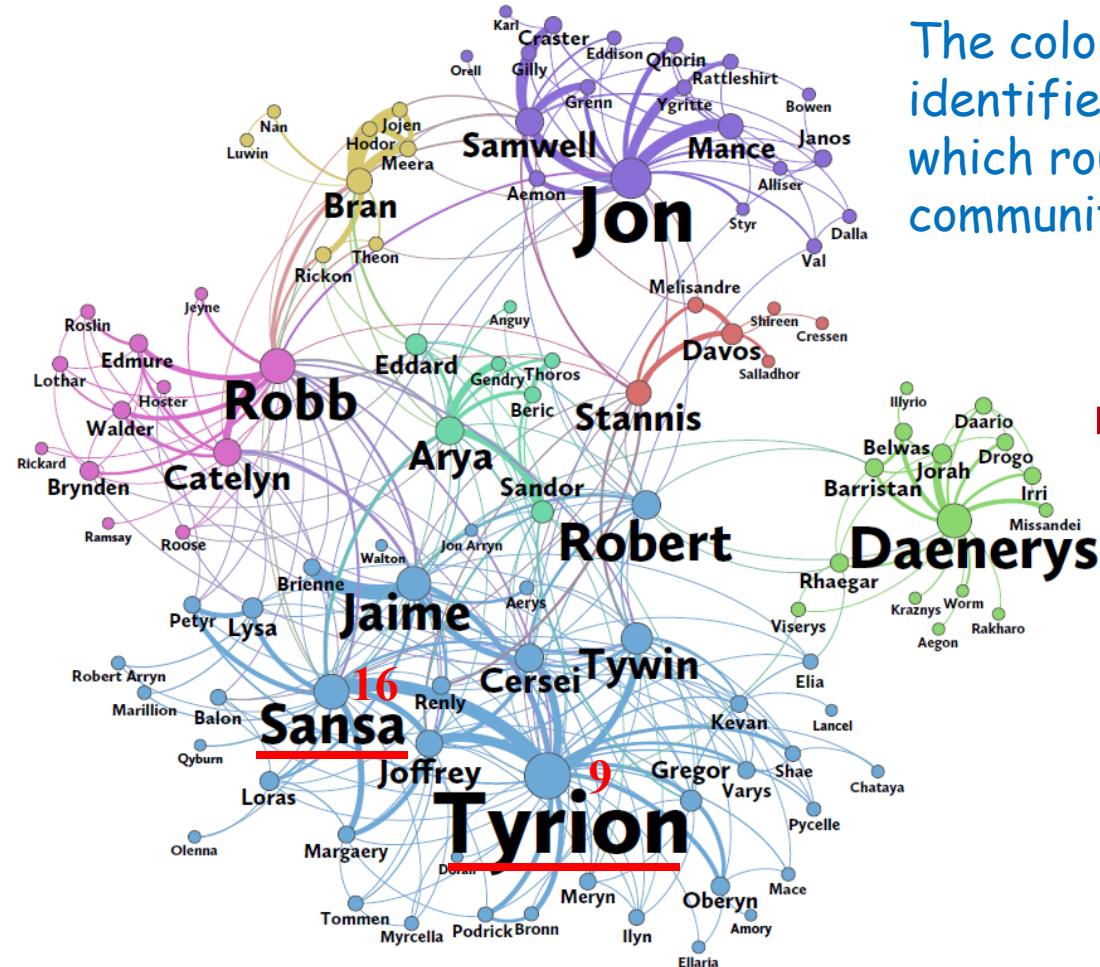


Illustration: Network of Thrones

Tie weights: the number of times the pair appeared within 15 words in the GoT book



The colors depict communities identified with network analytics, which roughly coincide with the communities in the book.

Ex. Tyrion Sansa

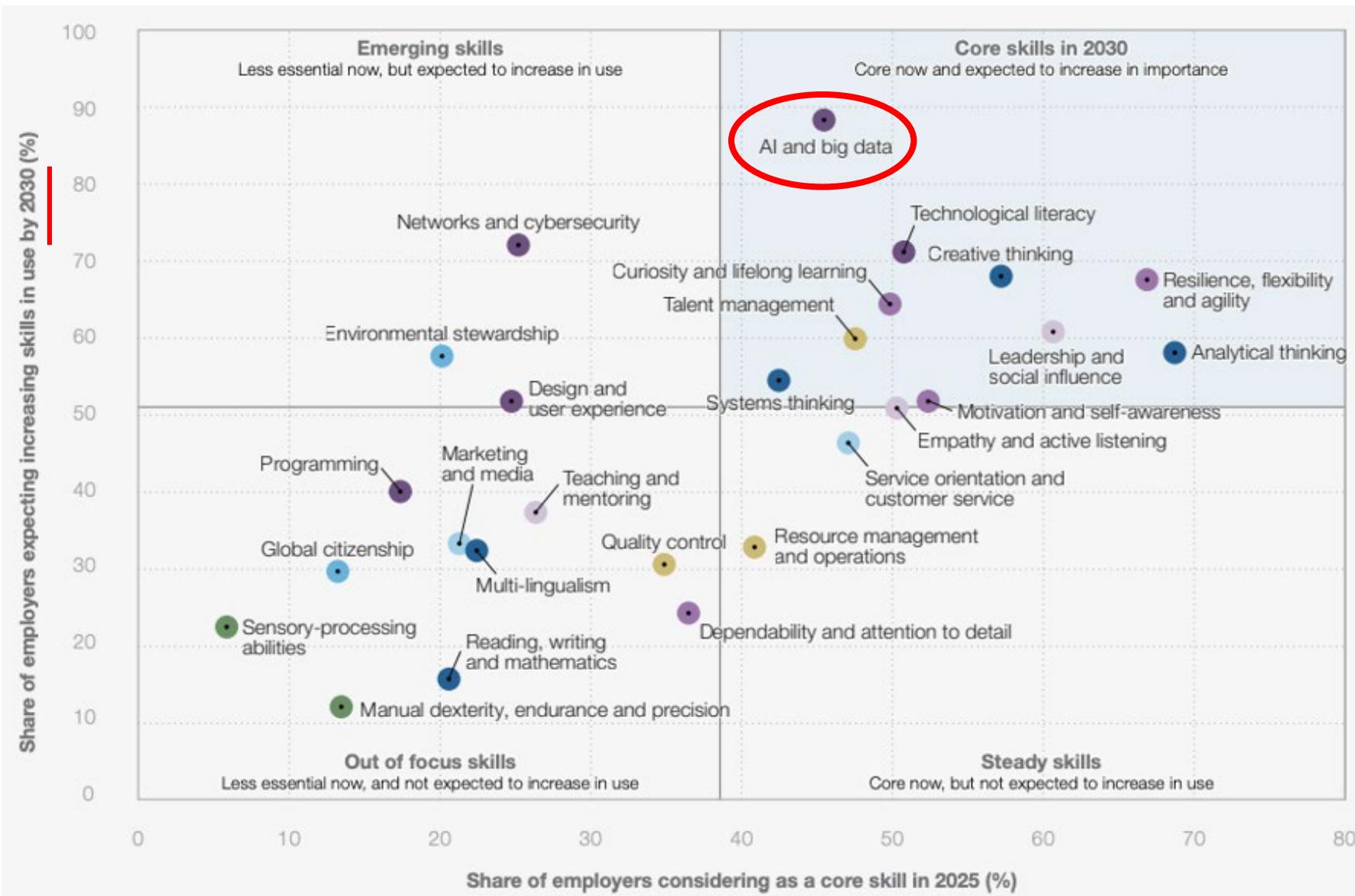
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|----|---|---|---|---|----|----|----|---|----|----|----|----|----|----|----|----|
| 1 | 0 | 0 | 3 | 8 | 8 | 8 | 3 | 8 | 8 | 3 | 0 | 0 | 3 | 3 | 0 | 3 |
| 2 | 0 | 0 | 3 | 8 | 8 | 8 | 3 | 8 | 8 | 8 | 3 | 0 | 0 | 3 | 3 | 0 |
| 3 | 0 | 0 | 3 | 8 | 8 | 8 | 3 | 8 | 8 | 8 | 3 | 0 | 0 | 3 | 3 | 0 |
| 4 | 3 | 3 | 3 | 3 | 3 | 3 | 0 | 3 | 3 | 3 | 0 | 3 | 3 | 3 | 0 | 3 |
| 5 | 8 | 8 | 8 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 8 | 8 | 11 | 11 |
| 6 | 8 | 8 | 8 | 3 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 8 | 8 | 11 | 11 |
| 7 | 8 | 8 | 8 | 3 | 0 | 0 | 3 | 0 | 0 | 0 | 3 | 8 | 8 | 11 | 11 | 0 |
| 8 | 3 | 3 | 3 | 0 | 3 | 3 | 3 | 3 | 3 | 3 | 0 | 3 | 3 | 3 | 0 | 3 |
| 9 | 8 | 8 | 8 | 3 | 0 | 0 | 0 | 3 | 0 | 0 | 3 | 8 | 8 | 11 | 11 | 0 |
| 10 | 8 | 8 | 8 | 3 | 0 | 0 | 0 | 3 | 0 | 0 | 3 | 8 | 8 | 11 | 11 | 0 |
| 11 | 8 | 8 | 8 | 3 | 0 | 0 | 0 | 3 | 0 | 0 | 3 | 8 | 8 | 11 | 11 | 0 |
| 12 | 3 | 3 | 3 | 0 | 3 | 3 | 3 | 0 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 13 | 0 | 0 | 0 | 3 | 8 | 8 | 8 | 3 | 8 | 8 | 8 | 3 | 0 | 3 | 3 | 0 |
| 14 | 0 | 0 | 0 | 3 | 8 | 8 | 8 | 3 | 8 | 8 | 8 | 3 | 0 | 3 | 3 | 0 |
| 15 | 3 | 3 | 3 | 8 | 11 | 11 | 11 | 8 | 11 | 11 | 11 | 8 | 3 | 3 | 3 | 0 |
| 16 | 3 | 3 | 3 | 8 | 11 | 11 | 11 | 8 | 11 | 11 | 11 | 8 | 3 | 3 | 0 | 0 |

Source: <https://networkofthrones.com>

Artificial Intelligence

Core skills in 2030

Share of employers considering skills to be a core skill in 2025 and share of employers expecting skills to increase in importance by 2030.



Source

World Economic Forum, Future of Jobs Survey 2024.

Note

The Future of Jobs Survey uses the World Economic Forum's Global Skills Taxonomy.
Bold lines represent the median values across all skills.

Artificial Intelligence (AI)

- Current consensus: **AI** will not replace jobs, people who know AI will !!
- **AI** is an **old** discipline – started in 1956 (Dartmouth Prof. John McCarthy)
- Defined as: "*the science and engineering of making intelligent machines that behave in ways that would be called intelligent if a human were so behaving*"
→ *Learning, Reasoning and Self-Correcting.*
- There are **many branches** of **AI**, such as:
 - ✓ **Logic-Based** – programs that follow rules, facts and human logic
 - ✓ **Expert Systems** – logic-based → decision rules elicited from experts
 - ✓ **Machine Learning (ML)** – *systems that learn mostly from historical data, without explicit programming*
 - ✓ **Generative AI (GAI)** – "*deep learning models that can generate high-quality text, images and other content based on data they were trained on*"
(IBM) → **generate** content from user **prompts**.

Current Trends in AI

- **Large Language Models:** generate human-like text, images, etc. to take the words you prompt and use **supervised learning** trained to predict the next word based on patterns it sees, over and over, sequentially.
- **Generative AI (GAI):** responds to prompts and generates responses as text, images, sound, video, etc.
- **Retrieval-Augmented Generation (RAG):** takes the GAI response and **augments** it with additional **factual** and **proprietary information**, improving quality and accuracy, and reducing **hallucination**.
- **Agentic AI:** Programming (with code or visually) a process of running a prompt (ad-hoc or automatic) through a GAI tool, using the output to feed as input to another (or the same) GAI tool, then again multiple times to accomplish a goal (e.g., booking a vacation)
- **Multimodal Models:** Can process and combine data from different sources (text, image, video, audio, etc.)

Superagency in the Workplace

McKinsey
& Company

- Survey of 3,613 employees and 238 C-level executives

<https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/superagency-in-the-workplace-empowering-people-to-unlock-ais-full-potential-at-work>

Superagency in the Workplace

i.e., human performance augmented by AI

Empowering people to unlock AI's full potential

- 92% of companies plan to invest in AI over the next 3 years
- 1% believe investments have reached maturity
- Employees are more ready for AI than their leaders are or imagine

Hannah Mayer
Lareina Yee
Michael Chui
Roger Roberts

January 2025



Prompt Engineering

It is the process of designing, refining, and optimizing prompts when interacting with AI models like GPT (Generative Pre-trained Transformer). Key aspects:

- **Clarity and Specificity** → ensure the model understands what's asked
- **Contextualization** → provide general context to get accurate response
- **Instructions for Tasks** → state what's expected (e.g., summary)
- **2 Methods:** (1) long, thorough prompt: (2) short prompt + iterations

Prompt:

"I am a machine learning professor at a business school. How do I explain what prompt engineering is to an audience of young students?"

ChatGPT Response:

"To explain prompt engineering to an audience of young students, you can draw parallels between how employees optimize business processes and how prompt engineers optimize AI outputs. Here's an example: Prompt engineering is the process of crafting precise, specific instructions for an AI model to ensure it provides accurate and relevant results, much like how you would design a financial query to get the correct data from a business database. In the same way that vague or poorly structured inputs can lead to incorrect human reports, unclear prompts can lead to less useful AI responses. By refining these prompts, we help the AI deliver exactly what we need, whether it's a financial summary, analysis, or any other task."

Prompt Engineering Example

Provided by an AI artist who creates images art in the style of known artists, describing what the image should contain along with key characteristics of the author's style.

Prompt (short version):

“In the muddy field, a man and a woman wearing traditional clothing stand near a wheelbarrow filled with hay. The man wears a brown jacket, while the woman is dressed in a mauve dress and a matching headscarf, seemingly in prayer. A picturesque backdrop features a landscape with a church, open fields, and a sky displaying hues of blue, orange, and purple, indicating either dawn or dusk. The pitchfork's tines are thicker.”

Generated with [ideogram.ai](#), by Arnold Millet



Prompt Engineering Example

Same prompt, done by me with ChatGPT, but adding “in the style of Jean-Francois Millet”

Prompt:

“In the muddy field, a man and a woman wearing traditional clothing stand near a wheelbarrow filled with hay. The man wears a brown jacket, while the woman is dressed in a mauve dress and a matching headscarf, seemingly in prayer. A picturesque backdrop features a landscape with a church, open fields, and a sky displaying hues of blue, orange, and purple, indicating either dawn or dusk. The pitchfork's tines are thicker. Render the image in the style of Jean-Francois Millet.”

Generated with ChatGPT, by me, same prompt



Prompt Engineering Example

Original famous painting: “The Angelus” by Jean-François Millet

How would you improve upon the previous prompt to render a more accurate drawing?



Final Words of Wisdom

- **Learn AI →** Take AI courses and/or enroll in AI programs; AI won't take your job, but someone who knows AI will !! AI is here to stay.
- **Learn to code →** Programming in Python & R is key.
- **The future →** Nobody can predict what you'll need to know in 10 years. What you'll learn will probably not be in use in 10 years, but it will help you understand the next generation of technologies.
- **Tech vs. Biz →** Figure out where your interests lie in the spectrum from **technical** (e.g., programming, data science, AI developer, etc.) to **business** (e.g., AI applications and uses, privacy, governance).
- **Personal use →** **Don't** use AI to do the **work for you** !! → If AI can do your work, you are not needed and AI can **hallucinate** → Use AI to **enhance** your work and productivity (e.g., writing, research, copilot).
- **Use responsibly and ethically →** there is great potential for unethical and harmful uses of AI → use AI for good and report unethical uses



Questions?



Reference Information



KOGOD SCHOOL *of* BUSINESS
AMERICAN UNIVERSITY • WASHINGTON, DC

Kogod School of Business @ American University

Information Systems, Analytics & AI Specializations & Major

Business Analytics and Artificial Intelligence Major (BAAIm)

(Major: 38 cr. Business Core + 12 cr. **BAAIm** Core & Programming + 15 cr. **BAAIm** Electives & AI Applications)

The major is designed to provide students with advanced technical skills and business knowledge in a data- and AI-driven world. Business analytics is about extracting meaningful insights from data to inform decision-making and drive AI. AI is the science of making machines perform tasks that typically require human intelligence, such as learning, reasoning, and problem-solving. This major integrates core business foundations with in-demand skills in data analytics, machine learning, and the responsible use of AI. Students acquire hands-on experience transforming raw data into actionable insights, building and applying AI models, and designing intelligent business solutions. BAAI affects most areas of business (e.g., sports, healthcare, marketing, etc.). The **BAAIm** program also provides students with a strong foundation for business analytics, data management, analytics modeling, and software tools (e.g., R, Python).

Business Analytics and AI Specialization (BAAIs) (Specialization: 10 cr. Core + 15 cr; Minor: 18 cr.)

The specialization is for Kogod BSBA students, and the minor is for non-Kogod students. **BAAI** affects most areas of business (e.g., sports, healthcare, marketing, etc.). The curriculum is similar to the **BAAIm**, but with fewer course requirements.

Information Systems and Technology Specialization (IST) (Specialization: 10 cr. Core + 12 cr; Minor: 18 cr)

The specialization is for Kogod BSBA students and the minor is for non-Kogod students. **IST** education focuses on learning how IT can be best utilized and managed to innovate products and processes, transform firms, and create business value. **IST** encompasses technology, people, business processes, and business models. **IST** professionals serve as a bridge between business and IT. Students who understand the functional areas (e.g., marketing, finance, accounting, HR, etc.) but also have a solid understanding of how IT works and how it can support business processes and models are ideally suited for **IST** careers such as consultants, business analysts, cybersecurity specialists, and IT project managers.

Legend: (RK) Required for all Kogod students; (GE) General electives; Business Analytics & Artificial Intelligence Major: (**BAAIm**) (bold) required; (BAAI) Electives; Business Analytics Specialization: (**BAAIs**) Required; (**BAAIs**) Electives; Information Systems & Technology Specialization: (**IST**) Required; (**IST**) Electives

| Core Courses (3 cr. each unless noted) | Pre-Requisites | Requirements |
|--|--|--------------------------|
| KSB 113 Navigating Business Technologies | None | RK |
| KSB 121 Data Analysis in Practice (2) | None | RK |
| KSB 224 Business Analytics | ITEC 200 or KSB 113 | RK |
| KSB 301 Introduction to Operations Management (2) | ITEC 200 or KSB 113 | RK |
| Required for Specializations & Major | | |
| ITEC 300 Programming Tools for Analytics (R) | ITEC200 or KSB113; STAT204 or KSB121 | BAAIm, BAAIs, IST |
| ITEC 321 Advanced Business Analytics with AI | ITEC 300; ITEC 320 or KSB 224 | BAAIm, BAAIs, IST |
| ITEC 455 Business Process Analysis | ITEC 200 or KSB 113 | BAAIm, BAAIs, IST |
| ITEC 470 Database and AI | ITEC 200 or KSB 113 | BAAIm, BAAIs, IST |
| ITEC 400 Programming Tools for Analytics (Python) | ITEC 200 or KSB 113; ITEC 300 | BAAIm, BAAIs, IST |
| Electives | | |
| ITEC 334 Online Business Development (Python) | ITEC 200 or KSB 113 | BAAIm, BAAIs, IST |
| ITEC 350 Digital Leadership | None | BAAIm, IST |
| ITEC 423 Social and Organizational Network Analytics | ITEC 200 or KSB 113; ITEC 300 | BAAIm, BAAIs |
| ITEC 460 Business Intelligence | ITEC 200 or KSB 113; STAT204 or KSB121 | BAAIm, IST, BAAIs |
| ITEC 466 Cybersecurity Risk Management | ITEC 200 or KSB 113 | IST |
| ITEC 467 Cybersecurity Governance | ITEC 200 or KSB 113 | IST |
| ITEC 468 Cybersecurity Analytics | ITEC 200 or KSB 113; STAT204 or KSB121 | BAAIm, IST, BAAIs |
| ITEC 477 Microsoft Certifications | ITEC 200 or KSB 113 | GE |
| DATA-312 Data Science Applications | | BAAIm |

BS/MS Dual Degree in Analytics and AI → You can **double count** up to 9 credits 6XX for 4XX, or 5XX:

ITEC 623 for ITEC 423; ITEC 630 for ITEC 455; ITEC 660 for ITEC 460; ITEC 670 for ITEC 470; ITEC 666 for ITEC 466; ITEC 667 for ITEC 467; ITEC 668 for ITEC 468; ITEC 5XX; Others as approved by MS Program Director

American University, Kogod MS Analytics

KSB 620 & 621 Capstone Experience Practicum (2 × 1.5 cr)

Electives & Functional Specialization (12 cr)

- ITEC 643 Project Management
- ITEC 630 Business Process Analysis
- ITEC 623 Organizational & Social Network Analytics
- ITEC 724 Advanced Text Analytics w/AI and Machine Learning
- ITEC 668 Cyber Security Analytics

Plus, specializations in:

- Cybersecurity
- Accounting Forensics
- Quantitative Financial Analysis I and II
- Marketing
- Accounting
- Data Science, etc.

Analytics Core (9 cr)

ITEC 610
Managerial Statistics

ITEC 620 Business Insights
Through Analytics

ITEC 621
Predictive Analytics
w/Machine Learning

ITEC 616
Managing the Digital
Organization

ITEC 670
Database & Big Data

ITEC 660
Business Intelligence

Tools:
KSB 999 R;
ITEC 600 Python

BS/MS Option → Double-count 9 credits

Further details: <https://jibe4fun.github.io/analytics/msanalytics.pdf>



A Few Useful General AI Tools

- **ChatGPT** (<https://chatgpt.com/>): conversational Generative Pre-Trained Transformer AI model. GPT-3.5 is free, ChatGPT Plus is \$20/month.
- **Perplexity** (<https://www.perplexity.ai/>): Conversational AI, similar to ChatGPT, but very accurate and yields useful citations
- **Copilot** (<https://www.microsoft.com/en-us/microsoft-copilot/personal-ai-assistant>): Microsoft AI tool with free and paid versions.
- **Gemini** (<https://gemini.google.com/app>): Google AI tool with free and paid versions.
- **Dall-E** (<https://openai.com/index/dall-e-3/>): generates graphs in ChatGPT
- **MidJourney** (<https://www.midjourney.com/home>): Generates images from text.
- **Pearl.ai, AI Code Editor, Claude, Cursor**: generative AI tools with advanced features for software programming.
- **Notebook LM** (<https://notebooklm.google.com/>): By Google, for generative AI on local documents like Word, Powerpoint and PDF.
- **Poe** (<https://poe.com/>): Great to run generative AI over local documents like Word, PowerPoint, and PDF files and render useful summaries.
See my Chatbot on Predictive Analytics: <https://poe.com/ITEC-621>
- **DeepL Translate** (<https://www.deepl.com/en/translator>): for accurate translation between multiple languages
- **Grammarly** (<https://www.grammarly.com/>): Excellent AI writing tool

Artificial Intelligence (AI) & Machine Learning Applications by Industry

- **Healthcare:** diagnostics, personalized treatment, robotic surgery, AI-driven diagnostic tools, drug discovery, predictive analytics for patient management.
- **Finance & Accounting:** risk management, fraud detection, accounting forensics, algorithmic trading, personalized banking services, and AI-driven compliance.
- **Retail:** personalized marketing, inventory management, sales forecasting, e-commerce, chatbots, recommendation engines.
- **Manufacturing:** production line optimization, predictive maintenance, quality control, automation.
- **Transportation:** self-driving cars, route optimization, logistics, traffic management, autonomous systems.
- **Energy:** energy use optimization, predicting equipment failure, managing grids, efficient energy consumption.
- **Advertising & Media:** content creation, targeted advertising, media analytics

Popular Uses of Generative AI

- **Research** → Search papers, summarize articles, find citations, etc.
- **Writing** → Write first drafts, edit writing, change styles (e.g., academic)
- **Local Knowledge Base** → Chat and generate from your own materials
- **Local Coach/Tutor** → Support learning and content creation
- **Advanced Searches** → Generate content, not just website links
- **Create Presentations** → Generate slides and graphics
- **Generate Creative Content** → Images, Music, Video, etc.
- **Software Programming** → Learn how to code, generate code, etc.
- **Language Translation** → Text or audio from/to any language
- **Power Computer User** → Tools (ex., Copilot for Excel, PowerPoint)
- **Voice Applications** → Voice assistants, transcriptions
- **Face Recognition** → Tagging photos, device security, etc.

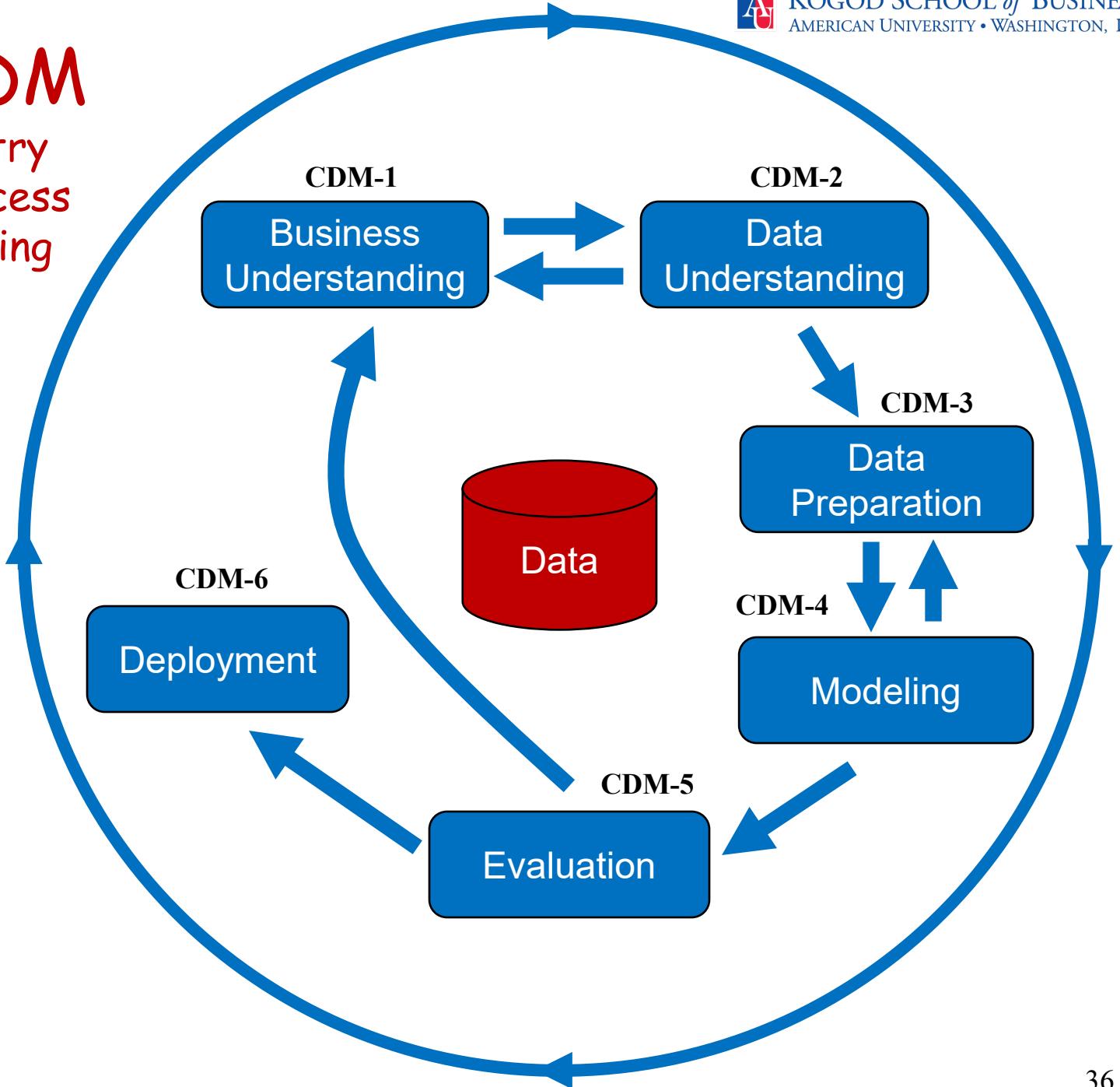
Backup Slides



KOGOD SCHOOL *of* BUSINESS
AMERICAN UNIVERSITY • WASHINGTON, DC

CRISP-DM

Cross-Industry
Standard Process
for Data Mining



Model Explainability

Interpretable Models

Sales ~ Ad Spending + Store Size

Explain how the model works and what the results **mean**
→ e.g., **OLS** regression, logistic regression, decision trees
→ good for **interpretability** and **inference**

Black Box Models (non-interpretable)

The complexities of the internal model are
difficult to explain

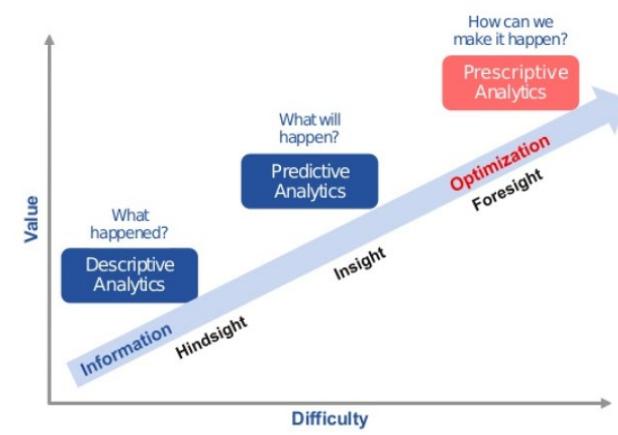
→ e.g., **neural networks**, deep learning
→ good for **predictive accuracy**



Stop?

Analytics Types

- **Descriptive:** getting meaning from the data – e.g., descriptive statistics, data mining, cluster analysis, market basket analysis, etc.
- **Predictive:** using some data to predict outcomes (quantitative or classification) – e.g., projecting sales/profits, decease survival prognosis, probability of losing/gaining a client, probability of a security breach, etc.
- **Prescriptive:** decision models that use predictive models to make/automate recommendations – e.g., optimal pricing, where to drill for oil, selecting job applicants, market products based on purchase history, etc.

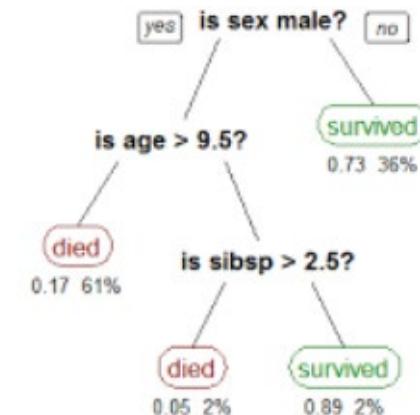
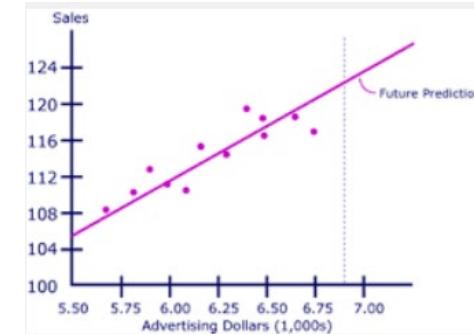


Derived from: Gartner (December 2012)

Predictive Analytics Question →

Outcome → What are you predicting?

- **Values:** → predict **quantitative** outcomes –
e.g., how much does annual income increase with each year of additional university education; how much sales increase for each additional dollar of advertising expenditures?
- **Classification:** predict **categorical** outcomes or the likelihood that an observation will fall in a particular class –
e.g., yes/no, positive/negative illness diagnosis, loan default/no-default, cyber security breach/no-breach, etc.



At Extras



KOGOD SCHOOL *of* BUSINESS
AMERICAN UNIVERSITY • WASHINGTON, DC

Moneyball book author: [Michael Lewis](#)

Billy Beane (Oakland A's) incorporated (“[sabermetrics](#)”) analytics and statistics, reaching record profits in baseball



Billy Beane

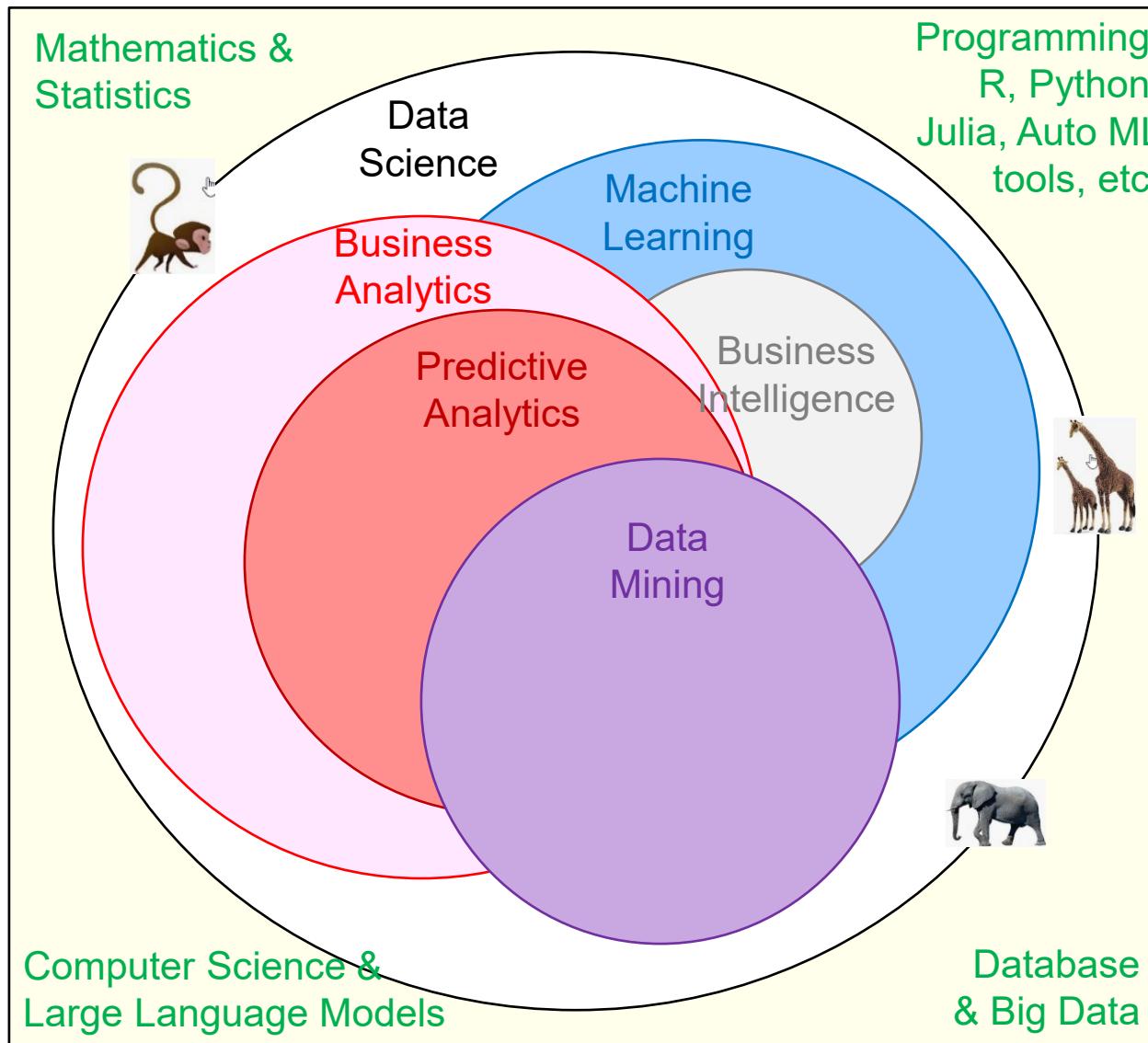
William Lamar Beane III (born March 29, 1962) is a former American professional baseball player and current front office executive. He is the executive vice president of baseball operations and minority owner of the Oakland Athletics of Major League Baseball. A first-round pick in the MLB draft by the Mets, Beane failed to meet the expectations of scouts, who projected him as a star. In his front-office career, Beane has applied statistical analysis (known as [sabermetrics](#)) to baseball, which has led teams to reconsider how they evaluate players. He is the subject of Michael Lewis's 2003 book on baseball economics, *Moneyball*, which was made into a 2011 film starring Brad Pitt as Beane.

Bill James

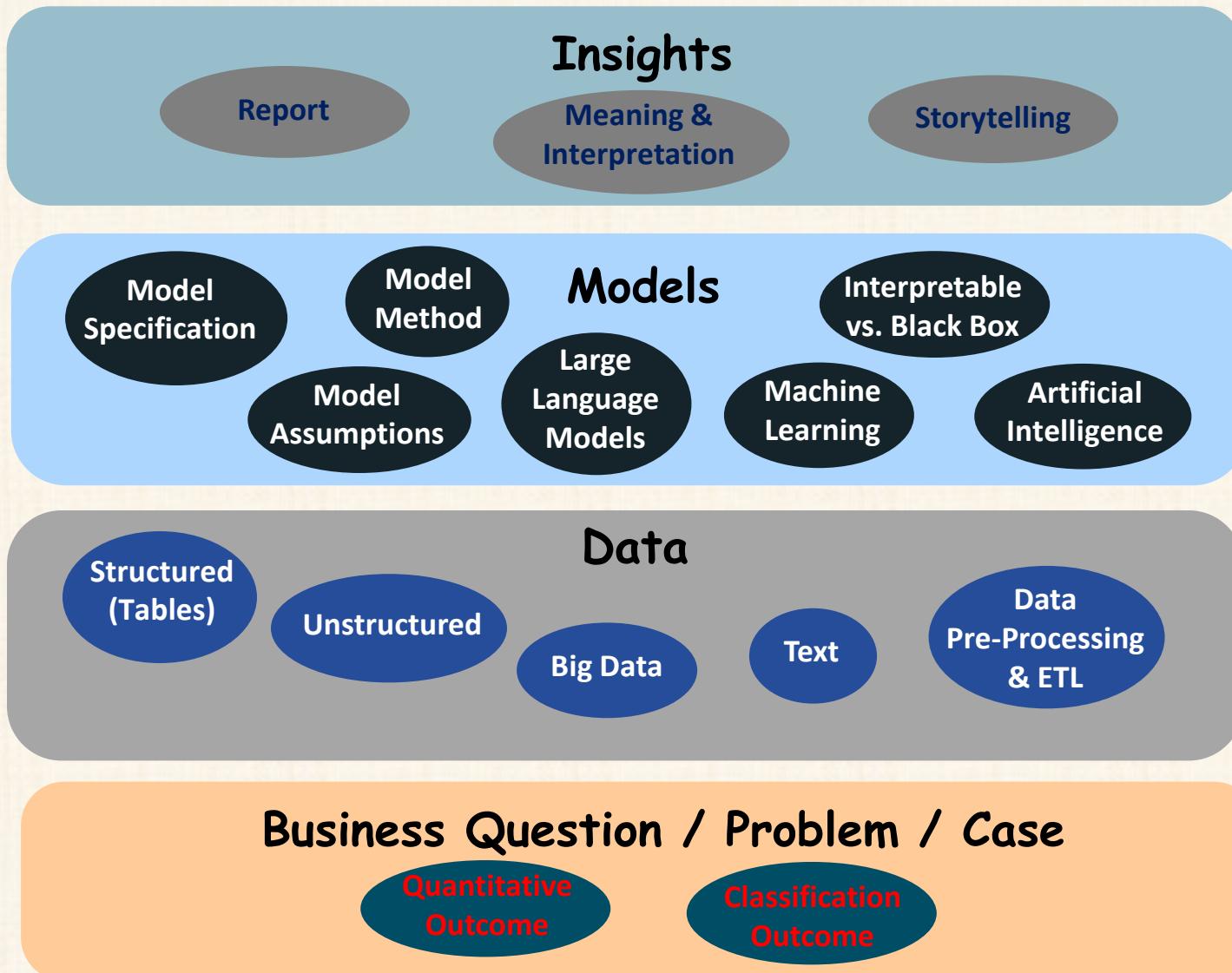


In 2006, *Time* named him in the *Time 100* as one of the most influential people in the world.^[3] He is a Senior Advisor on Baseball Operations for the Boston Red Sox. In 2010, Bill James was inducted into the Irish American Baseball Hall of Fame.^[4] His approach, which he termed [sabermetrics](#) in reference to the Society for American Baseball Research (SABR),^[1] scientifically analyzes and studies baseball, often through the use of statistical data, in an attempt to determine why teams win and lose.

AI & Data Science Zoo



Main Components of Analytics



The Analytics Life Cycle

Step 1 – Business Understanding (CDM - 1)

- (1) Formulate Business **Problem** and **Case**
- (2) Translate into **Analytics Question**: outcome + predictors

Step 2 – Data Work (CDM - 2, 3)

Identify & Gather data: structured, unstructured, visual, etc.

Pre-process data: cleanse, prepare, transform, format, etc.

Descriptive Analytics: e.g., correlation, descriptive stats, etc.

Step 3 – Modeling Method (CDM - 4)

Predictive Analytics → predict outcomes

Goals: Interpretability vs. Accuracy

Model Type: Explainable (e.g., regression) vs.
Black Box (e.g., neural network)

Prediction Type: Quantitative vs. Classification Outcome

Model Evaluation: Model assumptions, fit statistics,
cross validation testing

Model Specification: Predictors or features? Linear, non-linear, log?

Step 4 – Analysis (CDM - 5)

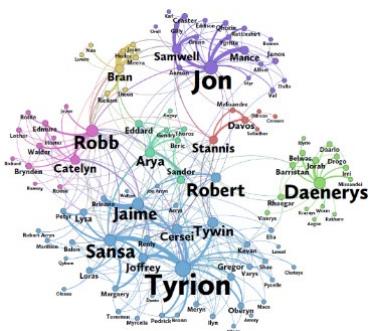
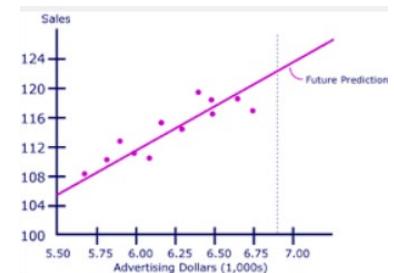
Interpret **effects**, test hypotheses and/or make predictions

Step 5 – Reporting (CDM - 6)

Written, interactive, visual → **"storytelling"**

Analytics Approaches

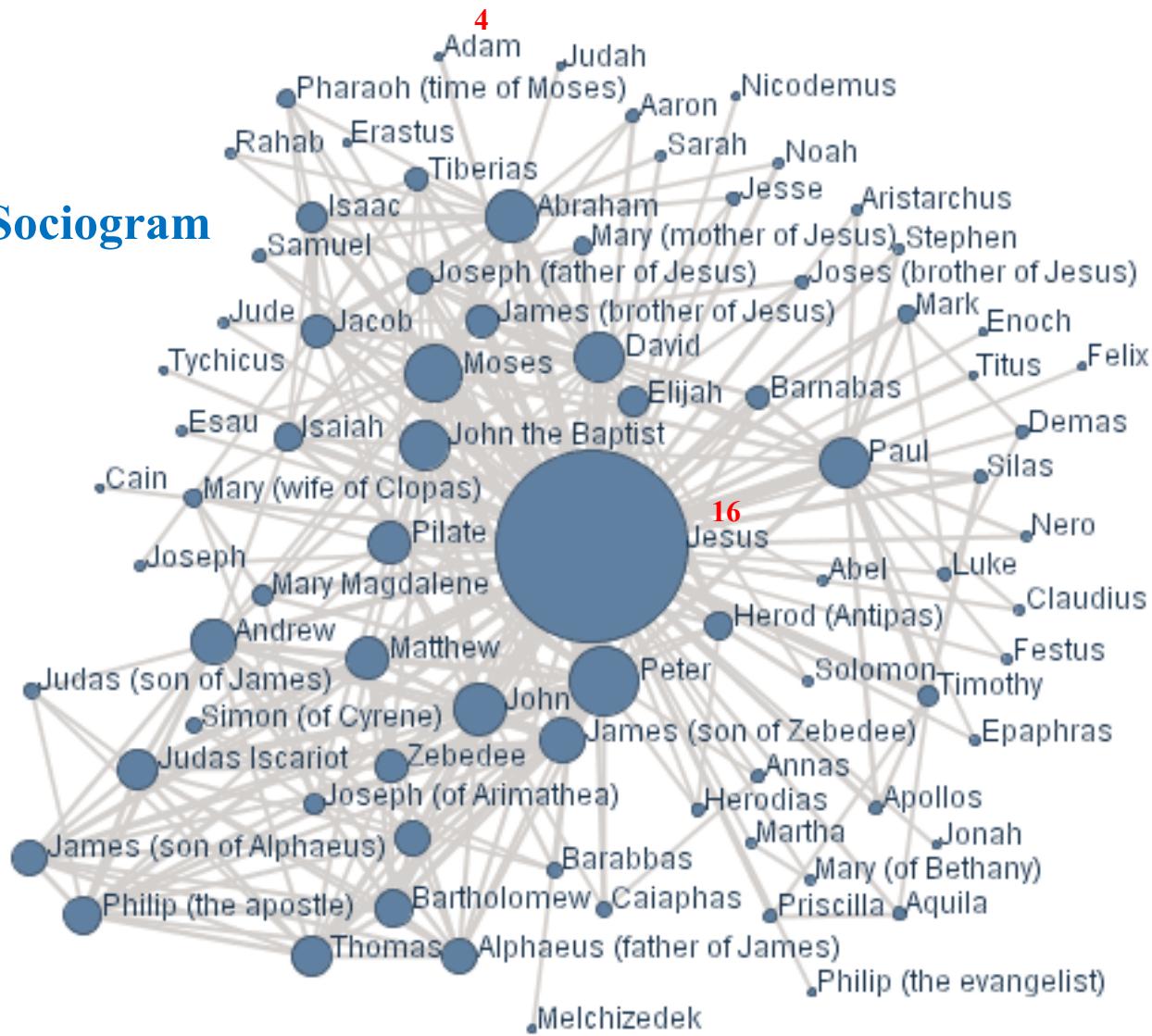
- **Quantitative:** statistical, association, correlation, regression, machine learning, etc. models based on how variables co-vary with the outcome – e.g., how much does annual income increase with each year of additional university education?
- **Visual:** e.g., Tableau, R, Python
- **Network Analytics:** relationships (e.g., friendship, communication, co-memberships, etc.) → social, organizational, professional networks, etc.
- **Unstructured Data Analysis:** text mining, natural language processing, social media, etc.
- **Large Language Models (LLMs):** predict the next word based on the current set of words in the text → generate text responses, e.g., ChatGPT



Network & Graph Analytics Illustration: The Bible (by IBM)

Tie weight: number of times the pair appeared in the same passage.

Sociogram



Ex. Adam Ex. Jesus

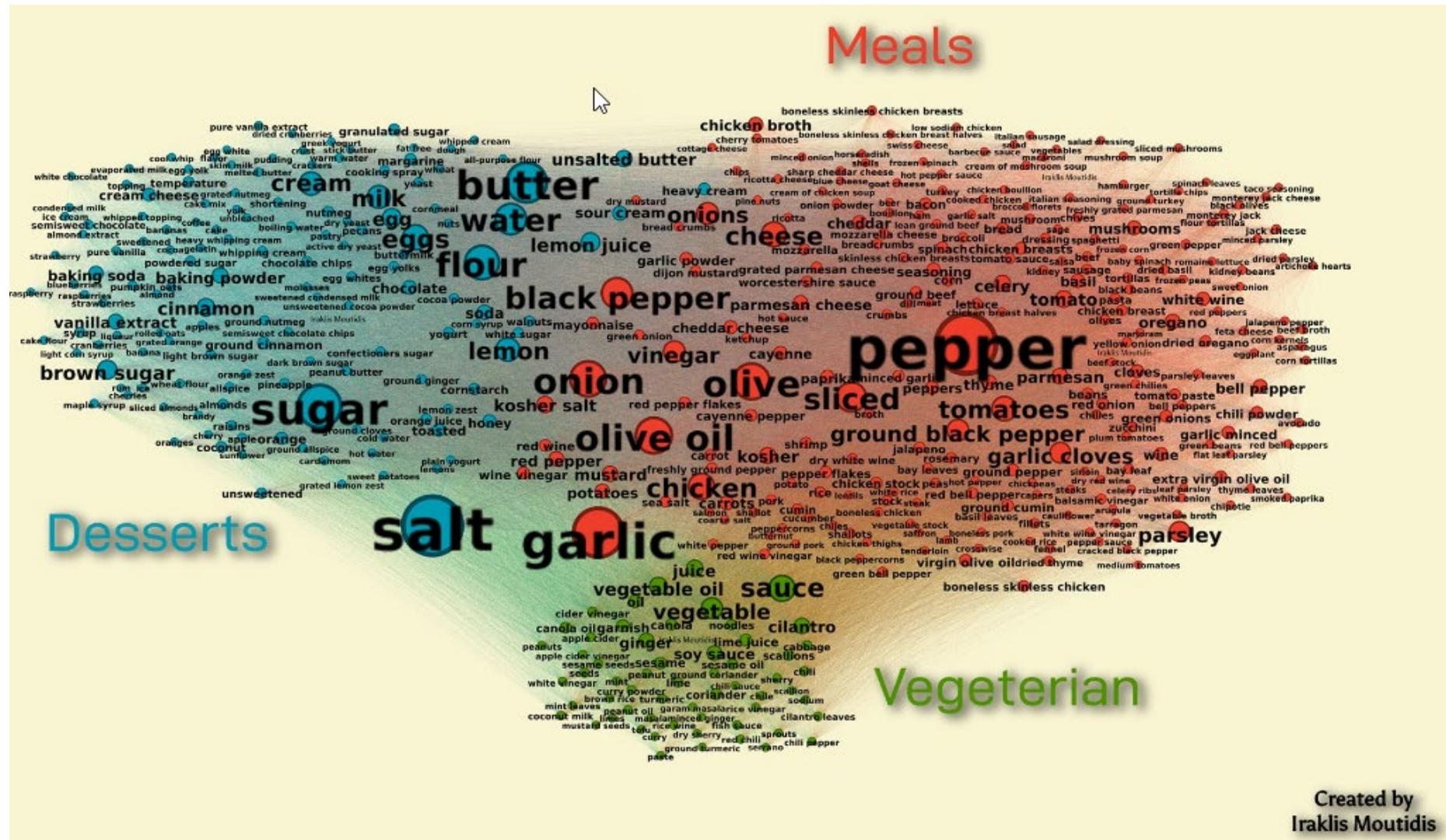
A 16x16 matrix representing the tie weights between 16 Biblical characters. The characters are numbered 1 through 16. Red arrows point from the labels "Ex. Adam" and "Ex. Jesus" to the first and last columns of the matrix, respectively. The matrix shows high values along the diagonal and some off-diagonal connections, such as between Adam (row 1) and Noah (column 10).

| | | | | | | | | | | | | | | | | |
|----|---|---|---|---|----|----|----|---|----|----|----|----|----|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | |
| 1 | 0 | 0 | 3 | 8 | 8 | 8 | 3 | 8 | 8 | 3 | 0 | 0 | 3 | 3 | 3 | |
| 2 | 0 | 0 | 3 | 8 | 8 | 8 | 3 | 8 | 8 | 3 | 0 | 0 | 3 | 3 | 3 | |
| 3 | 0 | 0 | 3 | 8 | 8 | 8 | 3 | 8 | 8 | 3 | 0 | 0 | 3 | 3 | 2 | |
| 4 | 3 | 3 | 3 | 3 | 3 | 3 | 0 | 3 | 3 | 3 | 3 | 3 | 3 | 8 | 8 | |
| 5 | 8 | 8 | 8 | 3 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 8 | 8 | 11 | 11 |
| 6 | 8 | 8 | 8 | 3 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 8 | 8 | 11 | 11 |
| 7 | 8 | 8 | 8 | 3 | 0 | 0 | 3 | 0 | 0 | 0 | 3 | 8 | 8 | 11 | 11 | 11 |
| 8 | 3 | 3 | 3 | 0 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 0 | 3 | 3 | 8 | 8 |
| 9 | 8 | 8 | 8 | 3 | 0 | 0 | 0 | 3 | 0 | 0 | 3 | 8 | 8 | 11 | 11 | 11 |
| 10 | 8 | 8 | 8 | 3 | 0 | 0 | 0 | 3 | 0 | 0 | 3 | 8 | 8 | 11 | 11 | 11 |
| 11 | 8 | 8 | 8 | 3 | 0 | 0 | 0 | 3 | 0 | 0 | 3 | 8 | 8 | 11 | 11 | 11 |
| 12 | 3 | 3 | 3 | 0 | 3 | 3 | 3 | 0 | 3 | 3 | 3 | 3 | 3 | 8 | 8 | 8 |
| 13 | 0 | 0 | 0 | 3 | 8 | 8 | 3 | 8 | 8 | 8 | 3 | 0 | 3 | 3 | 3 | 3 |
| 14 | 0 | 0 | 0 | 3 | 8 | 8 | 3 | 8 | 8 | 8 | 3 | 0 | 3 | 3 | 3 | 3 |
| 15 | 3 | 3 | 3 | 8 | 11 | 11 | 11 | 8 | 11 | 11 | 11 | 8 | 3 | 3 | 0 | 0 |
| 16 | 3 | 3 | 3 | 8 | 11 | 11 | 11 | 8 | 11 | 11 | 11 | 8 | 3 | 3 | 0 | 0 |

Sociomatrix

Illustration: Ingredient Co-Appearance Network

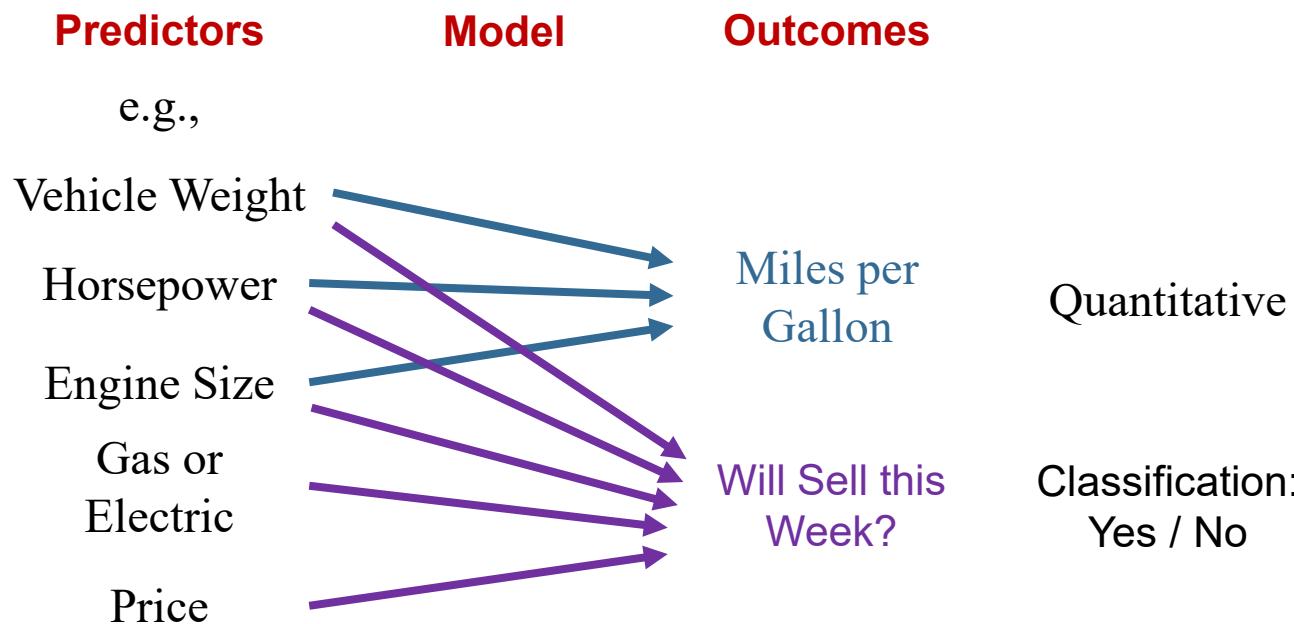
https://www.reddit.com/r/dataisbeautiful/comments/dor6l8/oc_ingredients_coappearance_in_one_million_recipes/



Created by
Iraklis Moutidis

Statistical Models

- **Regression** → predicting by **statistical association** using variables in the data to predict **quantitative** outcomes or **classifications**.
- **Tree** → partitioning the data at specific values of variables to differentiate **quantitative** outcomes or **classifications**



- Ordinary Least Squares
- Weighted Least Squares
- Generalized Linear Model
- Ridge Regression
- LASSO Regression
- Principal Components Regression
- Partial Least Squares
- Non-Linear Models
- Piecewise Linear Regression
- Step Regression
- Linear, Cubic, Smoothing Splines
- Neural Networks
- Etc.

Statistics Modeling Options

| Modeling Method | | | |
|--------------------|--|---|---|
| Predictive | Association | Decision Tree | Charts |
| Quantitative Value | Regression | Regression Trees | Regression plots, scatter plots, Tableau diagrams, trend charts, etc. |
| Classification | Logistic Regression; Other Categorical Regression Models | <ul style="list-style-type: none"> • Binomial Logistic Regression • Multinomial Logistic Regression • Linear Discriminant Analysis • Quadratic Discriminant Analysis • Neural Networks • Etc. | Tree maps, interactive diagrams, |
| Prescriptive | Operations research, optimization, linear | | lations, etc. |

ML: Key Concepts

- Again, **ML** is about **maximizing predictive accuracy** →
 1. **Split** the data set into **train** and **test** sub-samples
 2. **Train subset** → part of the data selected randomly to **fit models** and estimate parameters (e.g., regression coefficients)
 3. **Test subset** → part of the data, **not** in the **train** subset, used to **evaluate** the trained **models**
 4. **Cross-Validation (CV)** → Evaluate the **trained model's** predictive accuracy with the **test subset**, → CV Test Error
 5. **Re-Sample** and repeat 1-5 many times

Current Trends in Machine Learning

- **Large Language Models:** use supervised learning to predict the most likely words to add to text based on the input submitted.
- **AutoML:** Automate model selection, feature engineering, and parameter tuning.
- **Multimodal Models:** Can process and combine data from different sources (text, image, video, audio, etc.)
- **Foundation Models:** models like GPT that can be adapted to a wide range of applications.
- **Explainability for Complex Models:** Tools and methods are being developed to make models more transparent and to explain their predictions to non-experts.

Machine Learning

Cross-Validation → Split, Train, Test, Re-Sample

| Predictive Model Dataset | | | | | | | | | Outcome | | |
|--------------------------|------------|---------|-------|-------|-------|-------|------|--------|---------|-------|-----|
| Train | Sub-Sample | PizzaID | brand | mois | prot | fat | ash | sodium | carb | carb | cal |
| | | 14001 | D | 47.17 | 22.29 | 21.3 | 4.08 | 0.74 | 5.16 | 5.16 | 302 |
| | | 14002 | D | 49.16 | 27.99 | 17.49 | 3.29 | 0.39 | 2.07 | 2.07 | 278 |
| | | 14003 | A | 30.49 | 21.28 | 41.65 | 4.82 | 1.64 | 1.76 | 1.76 | 467 |
| | | 14004 | B | 52.68 | 14.38 | 25.72 | 3.26 | 0.93 | 3.96 | 3.96 | 305 |
| | | 14005 | H | 33.05 | 7.34 | 15.78 | 1.34 | 0.42 | 42.49 | 42.49 | 341 |
| | | 14006 | H | 35.55 | 7.32 | 16.4 | 1.76 | 0.36 | 38.97 | 38.97 | 341 |
| | | 14007 | G | 28.68 | 8.3 | 16.07 | 1.41 | 0.45 | 45.54 | 45.54 | 290 |
| | | | | | | | | | | | 51 |
| | | | | | | | | | | | -29 |
| | | | | | | | | | | | -11 |

Actual Predicted Error

Mean Squared Error: 1187.7

Mean Error = 34.5

Select Model

New Data

| | | | | | | |
|---|-------|-------|-------|------|------|------|
| D | 49.16 | 27.99 | 17.49 | 3.29 | 0.39 | 2.07 |
|---|-------|-------|-------|------|------|------|



Predict:

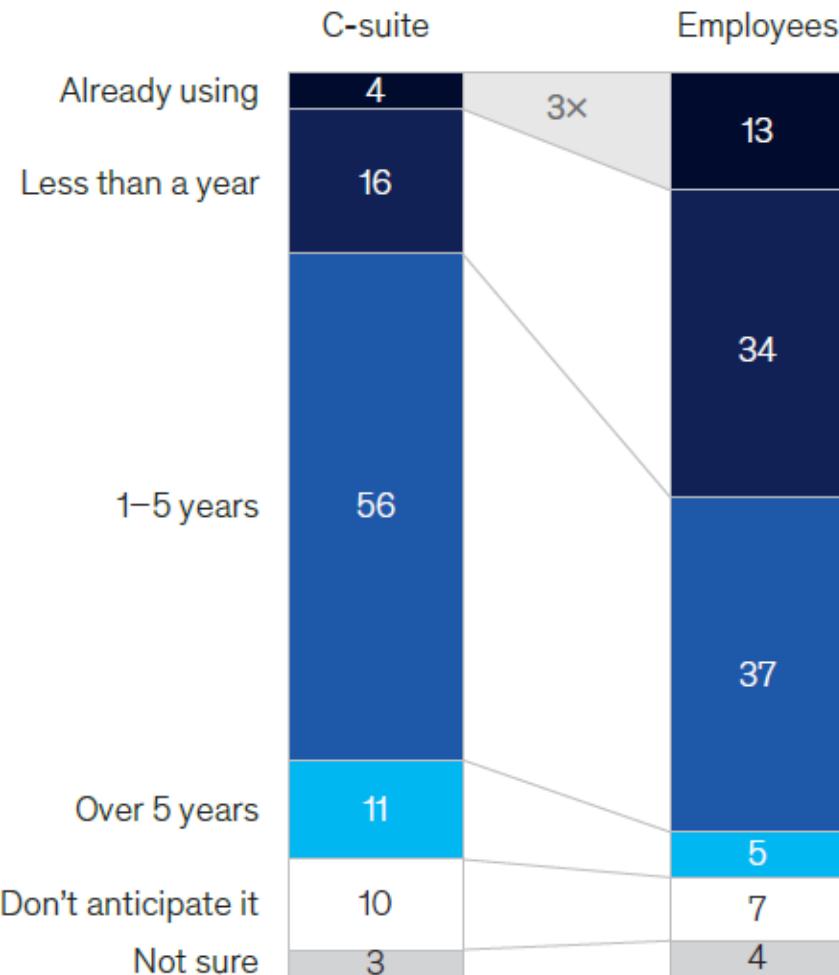
- Values
- Classifications
- Words
- Images, etc.

Uses: test and compare models, tune model parameters, train deep and supervise learning models, etc.

Artificial Intelligence (AI)

- Current consensus: **AI** will not replace jobs, people who know AI will !!
- **AI** is an **old** discipline – started in 1956 ([Dartmouth Prof. John McCarthy](#))
- Defined as: *“the science and engineering of making computers behave emulating human intelligence”* → learning, reasoning and self-correction.
 - *Gartner* → *AI is the application of advanced analysis, logic-based methods and machine learning (ML) to interpret events and automate decisions.*
- There are **many branches** of **AI**, such as:
 - ✓ **Logic-Based** – based on logical rules
 - ✓ **Expert Systems** – “if-then” decision rules are elicited from experts
 - ✓ **Machine Learning (ML)** – *systems that learn mostly from historical data, without explicit programming*
 - ✓ **Generative AI (GAI)** – *“deep learning models that can generate high-quality text, images and other content based on data they were trained on”* ([IBM](#)) → **generate** content from user **prompts**.

Employees vs. Leadership



Note: Figures may not sum to 100%, because of rounding.

Source: McKinsey US CxO survey, Oct–Nov 2024 (n = 118); McKinsey US employee survey, Oct–Nov 2024 (n = 3,002)

Current Trends in Artificial Intelligence

- **Generative AI (GAI):** ChatGPT and other tools that respond to a prompt and generate content based on it. The content can be text, images, videos, software code, etc. resulting in innovation and more efficient work processes.
- **Augmented Intelligence:** **copilot** decision makers, doctors, manufacturers, artists, etc.
- **AI in Predictive Analytics:** fraud detection, cybersecurity analysis, healthcare predictions, trend and financial forecasting trends, decision making, resource allocation and optimization.
- **AI in Personalization and Marketing:** personalize customers' experience, product and service recommendation engines.
- **Responsible AI and Ethics:** creating transparent, explainable and unbiased AI systems.

Agentic AI

- No or little human intervention → a programmed AI agent performs multiple tasks in sequence with various AI tools to achieve a goal
- It takes a prompt into one AI tool, passes the results to another AI tool, then another, etc.,
e.g., Ask ChatGPT → Send to Perplexity → Send to ChatGPT to Summarize → Output
- An AI agent can be programmed with tools like LangChain (requires Python skills) or visually with tools like Autogen Studio (free by Microsoft) and Flowise AI.
- Examples:
 - User inputs "**Schedule a call with team**" → AI Agent checks team members' calendars → Finds best times → Confirms with users → Schedules meeting and sends calendar invite
 - User inputs "**Find best wireless headphones**" → AI Agent searches Amazon, Reddit, and other review sites → Extracts top picks → Summarizes pros/cons → Output list with links
 - **Email received** → Agent reads and classifies intent (e.g., complaint, inquiry) → Drafts response using company policy + ChatGPT → Human approves → Agent Sends reply.