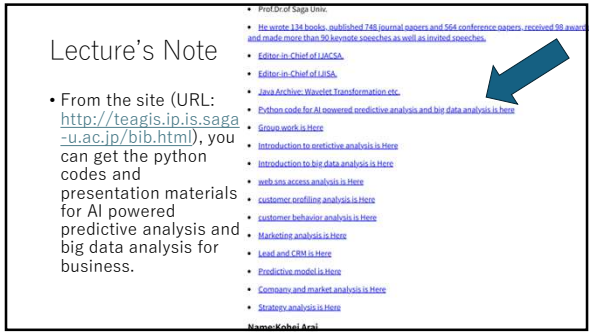
Data science, data analytics, predictive analysis, optimization, AI, and their transformative impact on the business world.

Kohei Arai
Saga University

Lecture's Note

- From the site (URL: http://teagis.ip.is.saga-u.ac.jp/bib.html), you can get the python codes and presentation materials for AI powered predictive analysis and big data analysis for business.



Oral explanation & Python Demo.

- 6 July: Introduction (Python code, Big Data Analysis, Predictive Analysis)
- 7 July: Data preparation (Data cleansing cleaning and EDA) ⇒ Python demonstration
- 8 July: Web access and behavior analysis ⇒ Python demonstration
- 9 July: Customer profiling, segmentation ⇒ Python demonstration
- 10 July: Customer value and satisfaction analysis ⇒ Python demonstration
- 13 July: Marketing effectiveness measurement and optimization ⇒ Python demonstration
- 14 July: Lead generation, nurturing, qualification and CRM ⇒ Python demonstration
- 15 July: Predictive analysis and modeling as well as machine learning ⇒ Python demonstration
- 16 July: Corporate and market analysis ⇒ Python demonstration
- 17 July: Wrap-up and Group Presentations (20 min. each by 4 groups = 80 min.) as well as report submission from every students

Group Works

Group 1

Mahmut	Sarı
Yeşim Nehir	Akgül
John	Braun
Arb	Lila
Melos	Hajrullahu

Group 2

Olti	Haxhiavdyli
Art	Haxholli
Blenda	Ponxhaj
Vera	Llugiqi
Genta	Llolluni

Group 3

Erza	Mustafa
Aulon	Krasniqi
Diellart	Mulolli
Sara	Sanwal
Manuel	Kabashi

Group 4

Atik	Hamiti
Arda	Ak
Rehan	Khaliq
Blesa	Dermaku
Majd	Abdeh

Report

- Report Assignment: Construct a multi-agent pipeline that proposes a use case where AI improves business operational efficiency and productivity.
- The project involves hypothesizing a dataset (or sourcing one from the web) to demonstrate the impact of the use case, performing necessary preprocessing—such as data collection and normalization—conducting analysis, visualizing the results, deriving insights, and proposing countermeasures.
- Finally, draft a report in the style of an academic paper that includes a discussion of these results.
- Submission Method: Email the completed report as a PDF attachment to arai@is.saga-u.ac.jp by 23:59 on July 16.

One of excellent group of PISU 2025 published the following paper on Computing Conference 2026

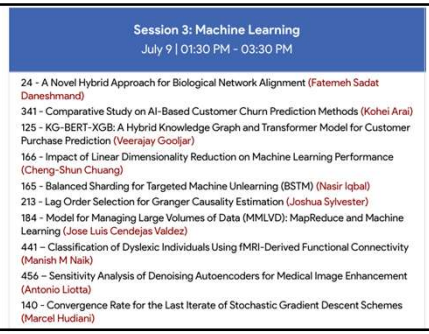
Comparative Study on AI-Based Customer Churn Prediction Methods to Provide Recommendations for Practical Implementation in Industry Settings

Kohei Arai1, Arbnor Pajaziti2, Fiona Beqa3, Tilda Tahiraj3, Rizone Hyseni3,Gabriel Khouri3, Diellart Mulolli3, and Jade Amponsah3

1Saga University, Saga city 840-8502, Japan

2 Department of Mechatronics, Faculty of Mechanical Engineering, Prishtina University

3 Former Students of Prishtina International Summer University 2025



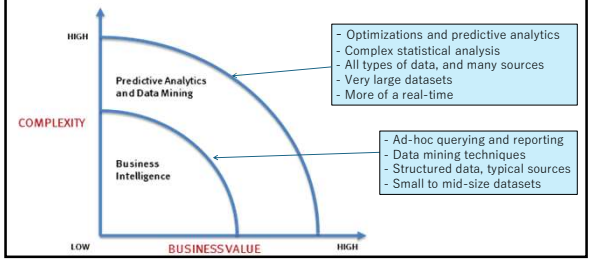
Purposes

- Key topics like **big data**, **predictive analytics**, and **machine learning**—especially as they apply to solving real-world **business challenges**.
- How organizations use data not just for technical purposes, but to **drive growth, improve efficiency, reduce risk, and create strategic advantages**.
- **AI models can analyze patterns in large datasets** to make accurate predictions—whether that's forecasting customer behavior, optimizing operations etc.
- DX, MCP-RAG for knowledge retrievals
- Agentic Engineering for data analysis pipelines (From data acquisition to insights and countermeasures)

Lecture 1 – Introduction to Big Data and Data-Driven Decisions:

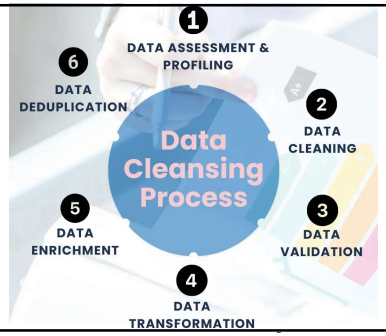
- We began by exploring the fundamentals of big data—massive datasets that exceed the capabilities of traditional tools due to their **volume**, **velocity**, and **variety**.
- These datasets are critical in a digital world where companies like Facebook, Walmart, and Airbus rely on real-time data to optimize operations.
- We discussed the shift toward **data-driven decision-making**, where insights derived from analytics and AI replace gut instincts, leading to faster and more effective business strategies.

What's driving Big Data



Lecture 2 – Data Preparation, Cleansing, and Exploratory Analysis:

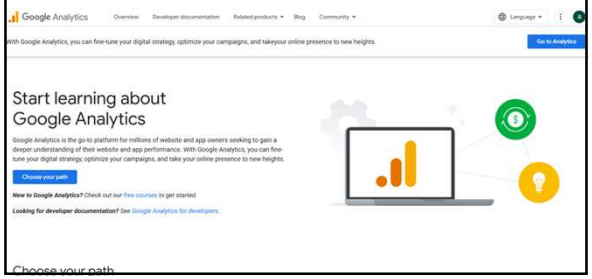
- This lecture focused on the vital first step of any analytics project: preparing and cleaning raw data.
- We distinguished between **data cleaning** (e.g., formatting, type conversion) and **data cleansing** (e.g., handling missing or inaccurate values).
- We also learned the importance of **Exploratory Data Analysis** (**EDA**) to uncover patterns, spot anomalies, and understand relationships using techniques such as **correlation matrices**, **histograms**, **scatter plots**, and **heatmaps**, supported by Python libraries like **pandas**, **NumPy**, and **seaborn**.



Lecture 3 – Web Access and Behavior Analysis:

- We analyzed how users interact with websites through metrics such as **store traffic**, **sales**, **marketing spend**, **satisfaction**, and **price**, and how these relate to one another through **correlation analysis**.
- We visualized these relationships using **heatmaps** and also introduced **Google Analytics API** for tracking real-time user data.
- This provided a framework to interpret customer behavior and evaluate marketing efforts based on live interaction data.

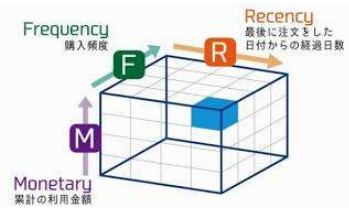
https://developers.google.com/analytics



Lecture 4 – Customer Understanding and Segmentation:

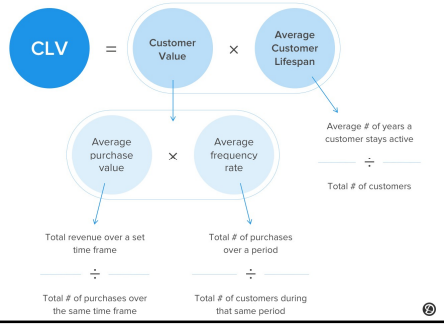
- This lecture introduced techniques for identifying and grouping customers based on behavior and value.
- We used **RFM Analysis** (Recency, Frequency, Monetary) to score and rank customer value, and **Decile Analysis** to divide them into performance-based tiers.
- We applied **K-Means Clustering** to segment customers into distinct groups using machine learning, and used **Cohort Analysis** to monitor customer behavior over time.
- These segmentation methods are crucial for targeting, personalization, and loyalty strategies.

RFM Segmentation



Lecture 5 – Customer Value and Satisfaction Analysis:

- We focused on estimating long-term customer impact through **Customer Lifetime Value (LTV)** and comparing it to **Customer Acquisition Cost (CAC)** to evaluate marketing profitability.
- Using **Customer Satisfaction (CS) Portfolio Analysis**, we assessed which aspects of service matter most to customers and where improvements are needed.
- We also introduced **Sentiment Analysis** with tools like VADER to interpret emotions in text from reviews and social media, giving businesses real-time feedback on public perception.



Lecture 6 – Marketing Effectiveness Measurement and Optimization:

- This session introduced strategic marketing frameworks and how to analyze them using data.
- **Funnel Analysis** helped us track user conversion across stages, identifying where customers drop off.
- **Conjoint Analysis** allowed us to assess how much customers value different product features.
- We also examined **4P (Product, Price, Place, Promotion)** and **STP (Segmentation, Targeting, Positioning)** models to align business strategy with customer needs.
- Finally, we used **ABC Analysis** to prioritize products or customers based on value tiers.

Lecture 7 – Lead Management and CRM Optimization:

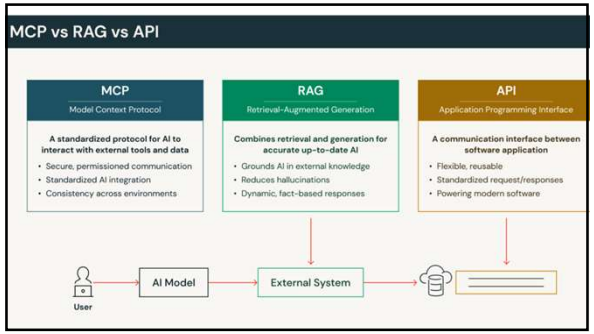
- We explored the full customer acquisition pipeline, from **lead generation** (via SEO, ads) to **lead nurturing** (email campaigns, webinars) and **lead qualification** (scoring interest and engagement).
- We introduced **Marketing Automation** and **Salesforce Automation (SFA)** to streamline repetitive tasks and manage customer interactions at scale.
- A lightweight CRM system was simulated using **SQLite**, and **Customer Journey Mapping** was used to visualize the path from awareness to support, helping improve the overall user experience.

Lecture 8 – Predictive Modeling and Machine Learning:

- This session introduced tools to forecast future outcomes using historical data.
- We began with **Linear and Nonlinear Regression** models such as **Lasso** and **Kernel Ridge**, then advanced to **Ensemble Learning** techniques like **Gradient Boosting**, **Random Forest**, and **XGBoost** for more accurate predictions.
- We applied these models to **visitor forecasting** and **churn prediction**, using the **Prophet** library for time-series analysis. These methods are essential for planning sales, marketing, staffing, and retention strategies.

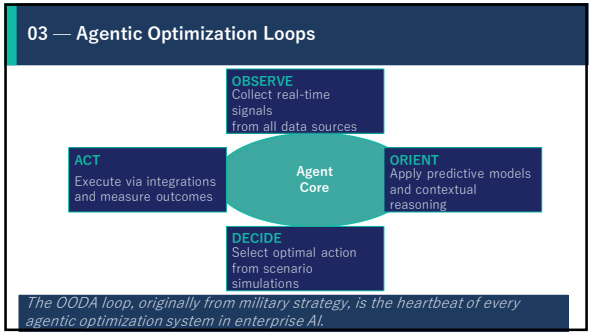
Lecture 9- DX for Business with MCP-Based RAG

- Definitions of DX, MCP, RAG
- Required stacks for DX, MCP-RAG
- **Business Use Cases**
- Sales Support and Customer Interaction: Consider a scenario where a sales representative consults the AI to prepare for a client visit next week: if MCP enables the AI to connect to internal systems, the AI can autonomously select and execute tasks—such as checking calendars, retrieving customer information, and searching for documents—based on the specific situation.



Lesson 10-Agentic Engineering-based Predictive Analysis and Optimization

- 1. What is agentic engineering?
- 2. Agentic AI workflows in business
- 3. Fundamentals of predictive analysis
- 4. Optimization for decision-making
- 5. Designing agentic pipelines for business impact
- 6. Case studies



Summary

- Traditional business intelligence-based approaches are less effective at streamlining operations and boosting productivity.
- Management decision-making driven by predictive analytics and optimization—leveraging big data through technologies such as data science, data analytics, and AI—is more effective for improving operational efficiency and productivity.