EAS 504: Applications of Data Science – Industrial Overview – Spring 2023

-Lecture by Matthew Nagowski and Eric Hanson

Name: Tananki Ranga Sai Saran Rohit

UB ID: 50441793

UB Email: rangasai@buffalo.edu

Ques 1 : Describe the market sector or sub-space covered in this lecture :

The market sector or sub-space covered in this lecture is Role of Data Science in Commercial banking. By providing liquidity to the financial system and guaranteeing that money are accessible to loan seekers even during economic downturns, the banking industry fosters economic stability. By providing liquidity to the financial system and guaranteeing that money are accessible to loan seekers even during economic downturns, the banking industry fosters economic stability. It helps people and companies to invest in their future and drive economic growth by facilitating financial intermediation between savers and borrowers. Data science is a multidisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data . data science plays a crucial role in helping commercial banking optimize their operations and improving customer experience. Data science may assist commercial banking in managing their loans , credit risks more efficiently and forecasting demand. Data science may be used to create financial models that assist businesses in forecasting cash flows, identifying possible hazards, and assessing investment possibilities. This entails processing massive amounts of financial data with statistical analysis and machine learning algorithms to produce prediction models.

Ques 2 : What data science related skills and technologies are commonly used in this sector?

Data science is growing rapidly and it is very important in the commercial banking sector, and now banking sector is recognizing the potential of data analytics to drive growth and profitability of the organization , also mitigating risks. With techniques to monitor, regulate, and manage massive amounts of data, data science has opened up a plethora of opportunities in the banking sector. Banks use data science to deliver the greatest service to their customers and to generate

future insights through prediction models and forecasting, risk management, compliance, safety, and so on. Machine learning is the application of algorithms to recognize patterns and make predictions based on data. Data analytics in banking has grown dramatically as the market's technical breakthroughs have increased. Furthermore, banks and the fintech industry give their clients with helpful and relevant data to forecast future positions and scenarios. Machine learning is used by commercial banks to create predictive models for credit risk assessment, fraud detection, and client segmentation. In risk management, SAS and Stata are widely used to evaluate and estimate credit risk, market risk, and operational risk. Banks utilize SAS and Stata to create models that forecast credit defaults, estimate default probability, and compute projected loss from loan portfolios. SAS and Stata are used to create models that mimic market scenarios and forecast future results. These models are used by banks to evaluate the influence of changes in interest rates, economic growth, and other factors on their business success.

Ques 3 : How are data and computing related methods used in typical workflows in this sector? Illustrate with an example.

Methods based on data and computation are crucial for identifying and combating fraud in the commercial banking industry. These approaches let banks to examine enormous amounts of data, discover patterns of fraudulent behavior, and take preemptive steps to avoid losses. Banks may safeguard themselves and their clients against financial loss by using data and computing-related approaches. Banks identify and prevent fraud using data and computing-related approaches. Banks, for example, may employ machine learning algorithms to examine enormous amounts of transactional data in order to uncover patterns of fraudulent behavior. Banks examine data using statistical analysis tools such as SAS or R to find trends and anomalies that may suggest fraudulent behavior. Dynamic Time Warping (DTW) is a time series analysis approach that compares and measures the similarity of two temporal sequences that may differ in length or speed. It is a non-parametric approach for aligning sequences in such a way that the difference between them is minimized. The DTW method generates a distance matrix representing the difference between each point in one sequence and each point in the other. The ideal path across the distance matrix that reduces the overall distinction among both sequences is then found. This is the best match between the two sequences. In risk management, DTW may be used to identify prospective threats and anticipate their impact on financial institutions. It is useful for analyzing historical data and identifying risk trends in financial markets, such as the link between default rates and interest rates. DTW can also be used to detect developing hazards, such as the impact of regulatory policy changes on financial institutions.

Ques 4 : What are the data science related challenges one might encounter in this domain?

The increasing need for data science related skills and technologies are key to the success of the retail technology sector to meet their increasing data size every day. They maintain large warehouses and inventory which leads to huge chunks of Data. Some of the data science challenges that can be encountered are data variability - different types of data , quantity - data size, speed - processing speed where data can often be complex and difficult to process and data can contain missing values, which are some of the obstacles that can be encountered in Commercial banking data when doing operations and analytics over it. One of the most difficult tasks is assuring data quality. Data might be partial, inconsistent, and erroneous, making proper analysis difficult. Moreover, data may be in numerous forms and kept in separate systems, making integration and analysis difficult. Automation and integration of data science technologies are crucial for managing massive amounts of data, while human bias reduction is critical for preventing biased effects. Human bias in data collection, analysis, and model creation may be a barrier for commercial banking data scientists. Biases can result in faulty forecasts, wrong insights, and discriminatory results. To guarantee fair and impartial decision-making, it is critical to recognize and reduce biases.

Ques 5 : What do you find interesting about the nature of data science opportunities in this domain?

Customer transactions, account information, financial markets data, risk management data, and other internal and external data sources all contribute to the massive volumes of data generated by commercial banks. According to a McKinsey & Company research, the total quantity of data created by the worldwide financial sector is predicted to exceed 44 zettabytes (44 trillion gigabytes) by 2020. This vast volume of data confronts commercial banks with both opportunities and problems, as they must be able to successfully gather, process, and analyze this data in order to acquire insights and drive business performance. Data science may assist by creating prediction models that detect possible fraud concerns before they arise. These models may examine a number of data sources, such as customer behavior, financial market data, and historical fraud trends, to detect possible fraud risks and enable institutions to take proactive fraud prevention actions. Data science may help commercial banks detect fraudsters by allowing banks to examine enormous volumes of data to identify trends and anomalies that may suggest fraudulent conduct. Data scientists may help banks avoid and identify fraudulent behavior in real-time by employing machine learning algorithms and sophisticated analytics, while also improving their fraud detection and prevention systems over time.

(i) According to the lecture, what are the types of technical and business questions that are considered to evaluate the validity of a model for a banking application?

The purpose of commercial banking model validation is to offer a comprehensive and systematic review of the models used by banks to make crucial decisions such as credit risk assessments, portfolio optimization, and capital planning. This helps to guarantee that banks use models that are accurate, dependable, and appropriate for their unique purposes, and that decisions are made using good mathematical and statistical concepts. The following are the questions that are asked during model validation: What other approaches were taken into consideration? Is the model stable from a conceptual standpoint? How accurate is the model in capturing the bank's portfolio's distinctive characteristics? representation of how the business world actually works in real scenarios? Is the selected methodology likely to be effective in light of the model's business purpose?

(ii) Describe how the clustering model meets the business purposes of M&T Bank, and what characteristics of the bank's portfolio were being captured.

The clustering approach consists of two steps, each with considerable control. The first is the estimation of distance, which contains 2 algorithmic modifications to make computing cost less. The overall time warp is constrained by the window size. In commercial banking like M&T Bank, the clustering approach is used to categorize client balances depending on their activity. The approach consists of two steps, the first of which is distance calculation, which contains algorithm enhancements to lower computation costs. The window size is adjusted using a variant of the DTW methodology, and the final dendrogram is created by clustering using a Hierarchical method that use Ward's D Statistic to minimize total within-cluster variance. The RPN count is calculated by aggregating the average behavior of 25 clusters into 5 broad behaviors, allowing the bank to understand balance behavior. Banks can identify RPNs with high balance potential by categorizing clusters depending on whether balances have risen or declined. This allows the bank to keep track of the characteristics of its portfolio, resulting in better informed decisions.