Assignment II: MIS 64038 Analytics in Practice

Case 2: Fraud Detection in Banking:

A community bank partnered with an analytics solution provider to develop a new fraud detection algorithm for ATM withdrawals. The bank provided historical data, and the company trained a model that seemed to provide acceptable performance when tested on the data. Once implemented, however, the bank faced a major tragedy: the algorithm was too slow in the production environment, and, as such, most ATM withdrawal requests were timed-out, and customers could not withdraw from their accounts. Discuss which aspects of the project were ignored and which phase(s) of the analytics project the problem can be associated to.?

Ans:

This case illustrates a common pitfall in analytics projects: the need for proper testing and validation of the model before deployment. The problem can be associated with two phases of the analytics project: model building and operations phase. Model building is the phase where the model's performance is assessed on unseen data, usually using metrics such as accuracy, precision, recall, etc. However, more than these metrics are needed to ensure the model works well in the production environment. Other aspects like scalability, robustness, latency, and compatibility should also be tested. For example, the model should be able to handle large volumes of data and requests without compromising its speed or accuracy. The model should also be compatible with the existing systems and infrastructure of the bank.

Model deployment is the phase where the model is integrated into the production environment and monitored for performance and impact. This phase requires careful planning and coordination between the analytics solution provider and the bank. The model should be deployed gradually and incrementally, with proper feedback mechanisms and contingency plans in case of failures or errors. The model should also be updated regularly to ensure its relevance and reliability.

To prevent this problem, the bank and the analytics solution provider could have conducted more thorough testing of the algorithm in a production environment. It would have allowed them to identify and address performance issues before implementing the algorithm. Additionally, they could have worked together to optimize the algorithm for the production environment to ensure it could handle real-world use demands. By taking these steps, they could have avoided the issue of slow performance and timed-out ATM withdrawal requests.

Abhinav Thupili- athupili@kent.edu (811234108)

Assignment II: MIS 64038 Analytics in Practice

Some best practices for implementing an analytics project include deciding on key metrics that are meaningful to your business[(1)], avoiding common data modeling mistakes[(1)], creating effective dashboards[(1)], and choosing the correct tool for your needs[(1)]. It is also important to take an analytics view of data, reconciling the questions the business asks with the kinds of data needed to deliver answers[(2)]. Building an effective analytics organization with deep functional expertise, strategic partnerships, and a clear center of gravity for organizing analytics talent can help ensure success[(3)].

Reference:

1. https://www.sisense.com/blog/4-ways-implement-data-analytics-best-practices/

2.https://mitsloan.mit.edu/ideas-made-to-matter/10-best-practices-analytics-success-including-3-you-cant-ignore

3.https://www.mckinsey.com/industries/financial-services/our-insights/building-an-effective-analytics-organization

Abhinav Thupili- athupili@kent.edu (811234108)

Assignment II: MIS 64038 Analytics in Practice

Case 3: Amazon Rekognition:

Amazon Rekognition is a cloud-based software as a service (SaaS) computer vision platform that was launched in 2016. It has been sold and used by a number of United States government agencies, including U.S. Immigration and Customs Enforcement (ICE) and Orlando, Florida police, as well as private entities. Rekognition provides a number of computer vision capabilities, which can be divided into two categories: Algorithms that are pre-trained on data collected by Amazon or its partners, and algorithms that a user can train on a custom dataset. In January 2019, MIT researchers published a peer-reviewed study asserting that Rekognition had more difficulty in identifying dark-skinned females than competitors such as IBM and Microsoft. In the study, Rekognition misidentified darker-skinned women as men 31% of the time, but made no mistakes for light-skinned men. The problem, AI researchers and engineers say, is that the vast sets of images the systems have been trained on skew heavily toward white men. In June 2020, Amazon announced it was implementing a one-year moratorium on police use of Rekognition, in response to the George Floyd protest. In May 2021, Amazon announced that they are extending its global ban on police use of its facial recognition software until further notice. Discuss which aspects of the project were ignored and which phase(s) of the analytics project, the problem can be associated to?

Ans:

The project of Amazon Rekognition has ignored some important aspects of data analysis, such as data quality, data diversity, model evaluation, and ethical implications. These aspects can be associated with different phases of the analytics project lifecycle, such as:

1. Discovery Phase.

2. Data Preparation Phase.

3. Model Building Phase.

4. Communications Phase.

- Data diversity refers to the representation and inclusion of different groups and categories of data in the analysis. It relates to the Discovery phase, which should be explored and analyzed to identify its features, patterns, and distributions. Data diversity can help ensure that the analysis is fair, robust, and generalizable to different scenarios and contexts. In the case of Rekognition, the data diversity might have been lacking by using data skewed heavily toward white men, which may not capture the characteristics and variations of other groups, such as dark-skinned women.

Assignment II: MIS 64038 Analytics in Practice

- Data quality refers to the accuracy, completeness, and consistency of the data used for analysis[(1)]. It relates to the data preparation phase, where the data should be cleaned, validated, and transformed to ensure quality. Data quality can lead to accurate and accurate results, as well as potential biases or errors in the analysis. In the case of Rekognition, the data quality might have been compromised by using data collected by Amazon or its partners, which may not reflect the diversity and variability of the real-world population.

- Ethical implications: This refers to the potential impacts and consequences of the analysis on individuals, groups, and society. It relates to the deployment and visualization phase, where the results should be communicated and presented honestly and responsibly. Ethical implications help ensure the analysis is aligned with moral principles and values, such as privacy, fairness, and accountability. In the case of Rekognition, the ethical implications might have been overlooked by selling and using the software for controversial purposes, such as immigration enforcement and police surveillance, which may raise ethical concerns and challenges.

- Model Building Phase refers to assessing and comparing different models or algorithms used for analysis. It relates to the model evaluation or building phase, where the models should be tested and validated using appropriate metrics and methods. Model evaluation can help to find the best model that fits the data and meets the objectives of the analysis. In the case of Rekognition, the model evaluation might have been insufficient by using algorithms that had more difficulty in identifying dark-skinned females than competitors, which may indicate a lower performance and accuracy of the model[(2)].

Reference:

1. https://blog.ppkn.co.id/5-bi-mistakes-and-how-to-avoid-them/

2. https://en.wikipedia.org/wiki/Amazon_Rekognition

Abhinav Thupili- athupili@kent.edu (811234108)

Assignment II: MIS 64038 Analytics in Practice

Case 4: IBM Watson in Healthcare:

Some time back, MD Anderson Cancer Center, the largest cancer center in the US, announced that it is going to introduce IBM Watson's computing system into the healthcare system. With the help of Artificial Intelligence, this system was supposed to accelerate the decision-making process of physicians while treating cancer tumors. But IBM Watson turned out to be a failure, as it did not deliver what it promised. It failed to analyze huge volumes of patients' health data and publish studies to offer cancer treatment options. Here are a few possible reasons why IBM Watson flopped in the healthcare industry, according to the experts. The AI technology that Watson uses is not a problem. The problem is that it is not given enough time to gather quality data and use personalized medicine. IBM launched Watson in a hurry to handle something as complex as healthcare. They were quite aggressive in marketing their product without realizing the importance of making it competent first. Watson was supposed to be launched as a software product in which oncologists can simply enter patient data and receive commendable treatment recommendations. This was how IBM advertised its Watson Health but it failed to deliver this effect. IBM failed to work with the hospitals to ensure the proper functioning of Watson. Another reason for Watson's failure is that IBM used data from its own development partner, MSKCC, to train it. Since the system is trained through the hospital's own data, the results it gave after queries were biased towards the hospital's own cancer treatments. It did not include data from other hospitals and other smaller clinical facilities. While such a trained system can be helpful in treating simple and generic cancer cases, complex ones need a different approach. Smaller hospitals cannot even access the same methods of treatment as their bigger counterparts. Discuss which aspects of the project were ignored and which phase(s) of the analytics project, the problem can be associated to?

Ans:

IBM Watson was a promising project that aimed to use artificial intelligence to improve cancer care. However, it failed to meet the expectations of the healthcare industry and the patients. These aspects of the project correspond to different phases of the analytics project lifecycle, such as:

- Data collection and preparation phase: It involves acquiring, cleaning, transforming, and integrating data from various sources and formats[(1)]. IBM Watson needed to collect and prepare more data from diverse and relevant sources to train its system. IBM Watson relied on data from a single hospital, MSKCC, to train its system. It resulted in a biased and limited dataset that did not reflect the diversity and complexity of cancer cases across different settings and populations. IBM Watson also needed more time to gather and analyze more data from other sources and validate its results.

- Model Planning Phase: It involves applying statistical and machine learning techniques to discover patterns, insights, and predictions from data. IBM Watson failed to analyze and model the data in a way that accounted for the complexity and variability of cancer cases and treatments.

- Communication Phase: This phase involves presenting and explaining the data analysis and modeling results in an understandable and actionable way. IBM Watson failed to visualize and communicate its results in a way that aligned with the oncologists' clinical workflows and decision support systems.IBM Watson did not involve the end-users, such as oncologists, nurses, patients, and administrators, in the design and development of its system. It needed to understand their needs, expectations, and challenges. It also could have communicated more effectively with them about the benefits and limitations of its system. IBM Watson also did not collaborate with other healthcare organizations and experts to ensure the reliability and validity of its system.

- Model building Phase: This phase involves implementing and testing the data-driven solutions in real-world settings and measuring their impact and outcomes. IBM Watson failed to deploy and evaluate its system in different healthcare settings and assess its effectiveness and efficiency.

References:

1.https://jrodthoughts.medium.com/some-ai-lessons-from-watsons-failure-at-md-anderson-9b895cf70840

2. https://thomaswdinsmore.com/2018/02/21/notes-on-a-watson-fail/

Assignment II: MIS 64038 Analytics in Practice

Case 5: AI for University Admission:

The researchers tried to develop a robot Todai, to crack the entrance test for the University of Tokyo. It's one of the tasks only humans can do with the required efficiency, but researchers thought they could train machines for this purpose. Unfortunately, the results exceeded their expectations, as AI was not smart enough to understand the questions. It would be better to introduce a broad spectrum of related information in the robotic system; so it can answer the questions rightly. Respective members from the National Institute of Information gave their statement about Todai: "It is not good at answering a type of question that requires the ability to grasp the meaning in a broad spectrum." Discuss which aspects of the project were ignored and which phase(s) of the analytics project the problem can be associated to.?

Ans:

The project of developing a robot Todai to crack the entrance test for the University of Tokyo was an ambitious attempt to demonstrate the capabilities of artificial intelligence[(1)]. However, the project failed to achieve its goal, as the robot needed help comprehending the questions that required a broad understanding of various topics and contexts.

One aspect that was ignored in the project was the diversity and complexity of human knowledge and reasoning. The entrance test for the University of Tokyo is not a simple test of factual recall or logical deduction but a test of critical thinking, creativity, and synthesis[(1)]. The questions often involve multiple disciplines, such as history, literature, science, and philosophy, and require connecting different ideas and perspectives. The robot Todai was trained on limited textbooks and past exam papers, which needed to capture the breadth and depth of human knowledge and reasoning. The robot Todai needed to have the common sense and contextual awareness humans have acquired through their life experiences and education.

Another aspect that was ignored in the project was evaluating and validating the robot's performance. The project focused on developing a system that could answer multiple-choice questions but needed to test how well the system could explain its answers or justify its reasoning. The project should have compared the robot's performance with human students or experts or solicited feedback from external reviewers or stakeholders. The project assumed that achieving a high score on the entrance test would be sufficient to prove the robot's intelligence but should have considered other aspects of intelligence, such as creativity, communication, collaboration, and ethics.

Assignment II: MIS 64038 Analytics in Practice

The problem of the robot Todai can be associated with several phases of the analytics project, such as data collection, data analysis, data interpretation, and data communication. In the data collection phase, the project needed more data to represent the diversity and complexity of human knowledge and reasoning. The project needed appropriate methods or algorithms to process and understand the data in the data analysis phase. In the data interpretation phase, the project did not validate or verify the results or outcomes of the data analysis. In the data communication phase, the project needed to communicate the results or outcomes clearly and convincingly.

In conclusion, developing a robot Todai to crack the entrance test for the University of Tokyo was a flawed and incomplete analytics project that ignored several important aspects of human intelligence and reasoning[(1)]. The project failed to achieve its goal because it needed to collect more data, analyze it properly, interpret it correctly, or communicate it effectively. The project also did not consider the ethical and social implications of creating a robot that could replace human students or teachers.

Reference:

1. https://thinkml.ai/five-biggest-failures-of-ai-projects-reason-to-fail/

2. https://nextshark.com/japanese-ai-robot-takes-university-entrance-exam-exposes-alarming-flaw-human-education

Abhinav Thupili- athupili@kent.edu (811234108)

Assignment II: MIS 64038 Analytics in Practice

Case 6: Mars Orbiter :

In 1999, NASA took a $125 million dollar hit due to the loss of a Mars orbiter. The loss was later attributed to a mix-up in the units of measurement used by Lockheed Martin's engineering team and NASA's internal team-Lockheed was using English units of measurement, and NASA was using more conventional metric system measurements. According to an internal review panel at NASA's Jet Propulsion Laboratory, "The loss of the orbiter was an end-to-end process problem... something went wrong in our system processes in checks and balances that we have that should have caught this and fixed it." Fixing this "end-to-end" process problem likely would have prevented this loss. NASA also blamed Congressional budget constraints for a portion of the error. So, additional funding would have also helped. Discuss which aspects of the project were ignored and which phase(s) of the analytics project the problem can be associated to.?

Ans:

The loss of the Mars orbiter in 1999 was a costly and embarrassing failure for NASA and Lockheed Martin. The root cause of the failure was a simple but fatal mismatch in the units of measurement used by the two teams involved in the project. Lockheed Martin's engineering team used English units of measurement, such as pounds and feet, while NASA's internal team used the metric system, such as newtons and meters. It resulted in a discrepancy in the orbiter's trajectory and velocity calculation, which led to its early entry into the Martian atmosphere and eventual disintegration.

The failure could have been prevented if the project had followed some basic principles of data analytics:

1. The project needed a clear and consistent data governance framework, which would have defined the standards, policies, and procedures for data collection, processing, and analysis. A data governance framework would have ensured that the data sources, formats, definitions, and units were aligned and consistent across the project teams.
2. The project did not perform adequate data quality checks and validations, which would have detected and corrected any errors or anomalies in the data. A data quality check would have verified that the data was accurate, complete, consistent, and reliable.
3. The project needed more data integration and transformation, which would have harmonized and standardized the data from different sources and systems.

A data integration and transformation process would have converted the data into a common format and unit of measurement that all the project teams could use.

The failure can be associated with several phases of the analytics project lifecycle. The most obvious phase is the data preparation phase, which involves collecting, cleaning, integrating, and transforming the data for analysis. The failure can also be traced back to the business discovery phase, which involves defining the project's objectives, scope, and requirements. The failure can also be linked to the Model building phase, which involves implementing and monitoring the analysis results. In each phase, the project teams needed to communicate more effectively, coordinate their activities, and adhere to best practices.

Reference:

1. https://degiuli.com/en/6-project-management-lessons-from-the-mars-climate-orbiter-failure/ - :~:text=In September 1999, the Mars,mission only to this problem.

2. https://www.simscale.com/blog/nasa-mars-climate-orbiter-metric/