# Assignment II: MIS 64038 Analytics in Practice

In this assignment, you will see a series of failed Business Analytics case studies and use cases, some of which are real-world examples. While they belong to different domains and disciplines, they have all failed one or more than one phases of the analytics project, hence the undesirable outcome. Your goal is to read through these uses cases and provide your insights as to why the project has failed. More specifically, you should discuss which phase(s) of the project (e.g., project discovery and definition, data accusation/preparing, model planning/building or operationalization) was inappropriately planned or executed (i.e., your diagnosis) and how the issue can be resolved (i.e., your proposed solution).

I have provided an example answer for Case 1 below and I leave Case 2-6 for you to answer.

## Case 1: Marketing Campaign

## This case represents a classic scenario of missing one-minute detail in the whole data mining process which ultimately led to a disaster scenario for a major Canadian bank. The case involved a logistic response model being built by an external supplier (not us) for acquisition of new customers regarding a given bank product of a well-known Canadian bank. The model was built and worked very well when looking at validation results. This model was then implemented and actioned on within a future marketing campaign. During the development process, the tools that were used both generated the solution as well as the validation results.

However, during the scoring process, the tool did not automatically generate the score. The user had to take the output equation results from the model development process and generate a scoring routine to score a given list of bank customers. In scoring, the user had to manually create the score by multiplying coefficients with variables. As part of this process, there was also a transformation of this equation to a logistic function. As part of this transformation, the user had to multiply the entire equation by -1. This fact of multiplying the equation by -1 was forgotten by the user when scoring the list of eligible customers. Guess what happened. Names with the highest scores represented the worst names with the opposite scenario happening for the lowest scores. The campaign went out by targeting names with the highest scores which ultimately resulted in horrific results.

When the supplier did the backend against a control random group of names promoted across all model deciles, they flipped the sign the right way to -1 and validated that the model worked quite well. Unfortunately, this did not appease the client's unhappiness as the bulk of their campaign names represented so-called targeted names within the top few deciles but who were in fact the worst names. From a net eligible universe of 500M names, the client ended up losing well in excess of $100M.

**Example Answer**

There are at least two areas (phases) of the project that could be linked to the failure of the above project: the commination phase and the operationalization phase. As for the communication phase (phase 5), it seems that the third party consulting firm failed to clearly communicate and/or emphasize enough on the importance of the steps needed to be taken to use the model’s output. Generally, once an analytics model is handed to a team, the development team is responsible to commutate the details of how the model should be used and to train the end users who will be interacting with the model.

In addition, the model did not seem to have been properly operationalized (phase 6), as it required a number manual operations which significantly increases the chance of errors. Automation should be used as much as possible to minimize the impact of human’s errors. In addition, checks and balances and models’ monitoring should be an integral part of the model’s operationalization. This scenario might have been prevented if there were checks and balances as part of the implementation process. By checking score distributions as well as the model variable means within the targeted deciles during model development and the current list implementation, this error would have been caught. The user would have noted that significant changes in both score distribution as well as model variable means for the targeted deciles would have occurred between time of model development and the current list scoring run. They then would have investigated this further by checking their coding in further detail and would have caught the omission and corrected it by multiplying the equation by -1. They say that the devil is in the details, but in data mining the devil is in the data.

Case 2: Fraud Detection in Banking

A community bank partnered with an analytics solution provider to develop new fraud detection algorithm for ATM withdrawals. The bank provided historical data and the company trained a model that seemed to provide an 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 were not able to 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?

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?

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 as something that can handle something as complex as healthcare. They were quite aggressive in the marketing of 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 their 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 to the 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?

Case 5: AI for University Admission

he researchers tried to develop a robot Todai, to crack the entrance test for the University of Tokyo. Its one of the tasks that only humans can do with required efficiency but researchers thought they could train machines for this purpose. Unfortunately, the results were opposite to their expectations as AI was not smart enough in understanding 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?

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, "IThe 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?