Module 4 XN Project: Project Roadmap

Esha Mulki

Mohammad Movahedi

Ajoy Kumar Nandakumar

Taiye Murtala

College of Professional Studies, Northeastern University

ALY6080: Integrated Experiential Learning

Dr. Matthew Goodwin

October 31, 2022

Abstract

This paper will outline the roadmap for the model creation for the sales forecasting project for the Danfoss sponsor. The document will include number of parameters starting with defining the analytical approach for the project. The machine learning algorithms and the tools and techniques used for the model creation. Performance of each of the model will be validated for the accuracy and then the best among them will be chosen for further analysis. The milestones to be followed to achieve success for the project is setup. A description of the job assignments for each of the team members will be defined. This will be followed by the key risk areas associated with the project and ways to mitigate them. Next, we define the metrics to measure the success of the project and lastly will discuss about the presentation method and delivery of the proof of concept.

Keywords: machine learning, project roadmap, sales forecasting

Analytic Approach

Sales forecasting is the practise of anticipating the volume of goods or services a sales unit will sell over the course of the upcoming week, month, quarter, or year in order to estimate future revenue(Kelwig et al.,2022). Inaccurate sales forecasting can lead to an undersupply or oversupply of inventory due to the failure in foreseeing the peaks or troughs in client demand, both of which can lead to huge loss for the company. An analytical approach is the understanding of business problem and expressing the problem in statistical and machine learning terms (Logallo et al., 2019). First step for any business problem is to analyze the data at hand completely. The initial analysis like data cleaning, reformatting, and exploratory analysis will be carried out using R or Python. If the data contains any missing values, decision needs to be taken about the ways to deal with it. Exploratory data analysis will be performed and various visualizations will be plotted. This will be done using the R or Python. Correlation between the variables will be checked to understand the variables most affecting the target variable. Also, multicollinearity between the variables will be determined and those variables will be dropped. Based on the findings from exploratory data analysis, feature engineering will be performed to understand the variables of interest for the model creation. The machine learning algorithm that will be chosen for the analysis are Time series analysis, K- nearest neighbours and Radom forest. The forecasted values would be validated with the actual data. RMSE and MAPE scores of the model would be calculated and the model with the least RMSE and MAPE values would be chosen for forecasting.

Milestones to measure progress

XN Project Roadmap
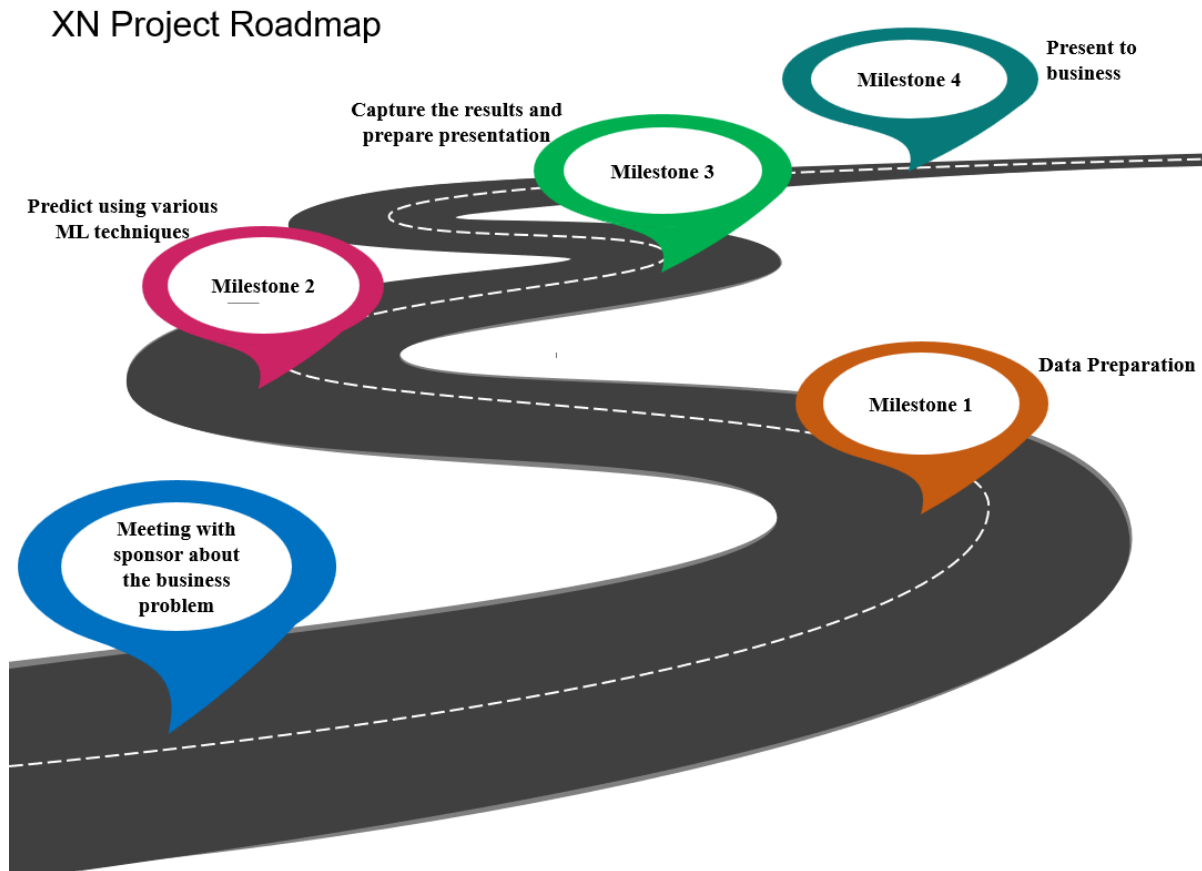
Fig 1: Milestones for the project

Milestone 1: Data Preparation

- Perform Exploratory Data Analysis on Data.

- Impute missing values

- Identify and implement any external variables if required

- Identify collinearity and remove multicollinearity

Milestone 2: Predict using various ML techniques

- ML techniques to be used are Linear Regression, Time Series analysis, k-Nearest Neighbours and Random Forest Regression

- Perform hyperparameter tuning

- Analyze the results of various ML techniques and choose the best model

Milestone 3: Capture the results and prepare presentation

- Document the results and analysis performed for the project.

- Create a story with evidence to be presented to the business.

- Review the presentation and finalise.

Milestone 4: Present to business

- Prepare points or insights to be highlighted to business.

- Prepare for questions.

- Deliver the presentation.

Job Assignment

Data cleaning and EDA – Team Member 1

Data modelling – Team Member 2

Dashboard design and Visualization– Team Member 3 and Team Member 4

Key risks and strategies to mitigate them

In the project, key risks can be found in three areas - designing the ML algorithm, executing the algorithm, and maintaining the algorithm. A machine learning algorithm creates a mathematical model after being fed data. Without having been explicitly coded, this model is capable of making a prediction and making a choice (GeeksforGeeks, 2021). In this process, the training

data is subjected to create risks such as bias and overfitting. We try to avoid these risks by analyzing every insight and comparing it with test data.

Security risks such as evasion and data poisoning attacks where malicious data inputs are provided can impact the classifiers and thus lead to inaccuracy of ML models (JavaT Point, n.d). Security risks can be avoided by retrieving data from authentic sources only and implementing data controls. The risks in executing and maintaining phases of the project include the risk of systematic error, malfunctioning, misuse, or ill intentions of final users and insights becoming irrelevant over time.

Measure of success

The model created using the machine learning algorithms will be tested for the accuracy and the RMSE and MAPE values. Model with better accuracy and least RMSE and MAPE values will be chosen for the prediction. The resultant model will be tested against test data to ensure there is no overfitting.

Presentation method and delivery of proof of concept

The final presentation will contain the insights and the analysis of the data provided. The results of the each of the models will be presented. RMSE and MAPE errors for each of the models will be given for the comparison. Finally, the conclusion based on the outputs of each of the models will be presented. The tools used will be the excel, R or Python, Tableau. The analysis will be visualized with the help of tableau dashboard for better understanding. The results of the model performance can be documented in the report for reviewing later.

References

Kelwig,D.(2022, June 24).*The definitive guide to sales forecasting methodologies.* Zendesk.

Retrieved from https://www.zendesk.com/blog/5-essential-sales-forecasting-techniques/

GeeksforGeeks. (2021). *Design a Learning System in Machine Learning*. GeeksforGeeks.

Retrieved from https://www.geeksforgeeks.org/design-a-learning-system-in-machine-learning/

Java T Point. (n.d.). *Risks of Machine Learning*. https://www.javatpoint.com/risks-of-machine-learning

Logallo, N. (2019, December). *Data Science Methodology 101*. Towards Data

Science. https://towardsdatascience.com/data-science-methodology-101-ce9f0d660336