



**College of Professional Studies
Northeastern University San Jose**

MPS Analytics

Course: ALY 6080

Assignment:
PROJECT ROADMAP

Submitted on:
May 25, 2023

Submitted to:

Professor: Venkata Duvvuri

Submitted by:

Archit Barua

Nikshita Ranganathan

Heejae Roh

Shyamala Venkatakrishna

Your analytic approach to answer the project requirements (what models/tools/techniques)

To meet the project requirements, our approach will incorporate several key methodologies and tools that will streamline the problem definition, modeling, solution development, and result presentation. The underpinning of our work will be grounded in the robust analytics that pave the way to valuable insights.

We'll employ Python for data cleansing, manipulation, and model creation. Python's extensive range of libraries and frameworks facilitates efficient data handling and offers a plethora of machine learning algorithms to choose from, catering to our diverse needs.

Excel will play a crucial role in our preliminary data exploration and understanding, enabling us to delve deeper into the data set and identify potential patterns or anomalies.

Lastly, Tableau will be the tool of our choice for data visualization, which will assist in making our findings more understandable and visually appealing. It will be key to transforming our complex data into easily interpretable visual representations, thus aiding in the presentation delivery.

The culmination of this process will be a comprehensive analytic solution tailored specifically for the project sponsor, comprising of interactive dashboards and sophisticated models, thereby ensuring the deliverables' precision and usefulness.

Milestones to measure progress

- **Initial Project Setup:** This stage involves familiarizing ourselves with the individual skill sets of the team and finalizing the Non-Disclosure Agreement (NDA) to maintain the confidentiality of proprietary data.
- **Data Acquisition and Initial Exploration:** This phase includes the tasks of loading, merging, and performing initial statistical characterization of the data, providing us with an understanding of its nature and structure.
- **Data Evaluation and Preliminary Model Development:** During this stage, we'll scrutinize the datasets and initiate the model training and validation process, ensuring the adequacy of the models chosen.
- **Model Targeting - Building Customer Email Look-alike:** Here, we'll focus on constructing a specific model to generate look-alike customer email profiles. For this purpose, we'll employ K-modes clustering, a technique that's particularly effective for categorizing categorical data, such as email characteristics.

- **Applying Models to Address Business Objectives:** In this pivotal phase, we'll apply the developed models to answer business-specific objectives and questions. The insights drawn here will be instrumental in facilitating strategic decision-making and problem-solving in the business context.
- **Visualization of Analysis Results:** This phase involves transforming the statistical analysis into visual forms like charts and diagrams using Tableau. These visualizations make it much simpler to interpret data trends and patterns.
- **Impact Analysis:** This stage involves thorough research on appropriate algorithms and a detailed evaluation of the project's alignment with the business objectives. We'll determine how our findings and model outcomes could potentially influence business goals.
- **Quality Assurance and Revision:** The final milestone encompasses a comprehensive check and revision of all the work done. We'll confirm the accuracy and relevance of our work, make necessary adjustments, and ensure that all aspects perfectly align with the project requirements.

Job assignments of each group member

Each group member has been assigned specific tasks that align with their skills and the project's needs:

Heejae and Shyamala: They focused on the Exploratory Data Analysis (EDA) and the crucial task of model building. Their work involved understanding the underlying patterns in the data, defining relationships between variables, and constructing the K-modes clustering model. They played a significant role in shaping the analytics foundation of the project.

Archit and Nikshita: They were assigned the responsibility of conducting comprehensive company research and providing a complete business overview. Their work also involved a detailed understanding of the business objectives. Their insights helped align the technical work with the business perspective, ensuring that the project remains relevant and beneficial to the company's strategic goals.

Key risks and strategies to mitigate them

- In any project, there are potential risks that could hinder progress or affect the quality of outcomes. Here are some key risks that we may face in our project, along with strategies to mitigate them:
- **Data Quality:** Poor data quality is a significant risk that could adversely affect the entire project. To mitigate this, we'll ensure to thoroughly clean and preprocess the data, handle missing values, and check for consistency and validity before conducting any analysis.
- **Model Inaccuracy:** The risk of building an inaccurate model can result in flawed predictions or interpretations. To mitigate this, we'll use cross-validation techniques and fine-tune model parameters to optimize performance. Regular testing and validation will also ensure that our models remain accurate and robust.
- **Overfitting:** This occurs when the model learns the training data too well and performs poorly on unseen data. We'll mitigate this by using techniques like regularization, using a validation set, and early stopping.
- **Team Coordination:** Miscommunication or misunderstanding between team members could cause delays or errors. Regular meetings, clear communication, and well-defined roles and responsibilities can help mitigate this risk.
- **Changing Business Objectives:** The company's objectives might change during the course of the project, which could render our results less useful. Regular communication with company stakeholders and an agile approach to the project can help us adapt to changes swiftly and efficiently.
- **Security Breach:** There's a risk of sensitive data being compromised during the project. To mitigate this, we'll adhere to strict data security protocols, use secure systems for data storage and processing, and ensure all team members understand and follow best practices in data security.

These strategies, among others, will be instrumental in ensuring that the project runs smoothly and successfully, despite potential challenges.

Measure of success

The success of our project can be measured using several key metrics:

- **Model Performance:** The predictive or clustering accuracy of our model, as indicated by appropriate metrics (such as precision, recall, F1-score for classification tasks, or silhouette score for clustering tasks), will be a direct measure of success.
- **Alignment with Business Objectives:** Our success will be judged on how well our outputs and recommendations align with the company's business objectives. This involves both solving the defined problem and contributing to the overall growth strategy.
- **User Acceptance:** The adoption and acceptance of our deliverables (like dashboards, reports, or models) by the end-users within the company will also be a key measure of success.
- **Value Created:** Our project's success can also be evaluated based on the value it creates for the business. This could be in terms of cost savings, increased revenue, improved decision-making, or other business benefits.
- **Timely Delivery:** Completing the project within the stipulated timeline while maintaining the quality of deliverables is another important measure of success.
- **Risk Mitigation:** Successfully identifying potential risks and mitigating them effectively without causing major delays or issues in the project will be another indicator of successful project management.

By tracking these metrics, we can effectively measure the success of our project, ensure we're meeting our goals, and make necessary adjustments if needed.

Your presentation method and delivery of your proof of concept

Our presentation method and delivery of the proof of concept (PoC) will be both engaging and interactive, designed to effectively communicate our findings and illustrate our approach. Here's how we plan to execute it:

- **Structure:** We'll begin with an overview of the project, including its objectives, scope, and key milestones. Following this, we'll delve into our methods, discussing our data exploration, model development, and analysis in detail. We'll then present our key findings, showing how they address the business objectives, and finally conclude with next steps and potential improvements.
- **Visualizations:** We will use Tableau to create clear, compelling visualizations that illustrate our findings and the insights derived from our analysis. Graphs, charts, and other visual tools will be used to depict trends, patterns, and relationships in the data.
- **Live Demonstration:** We will provide a live demonstration of our PoC, showcasing how our model works and the kind of results it generates. We'll also demonstrate any dashboards or tools we've developed as part of our project.
- **Clear Communication:** We will use clear, non-technical language to ensure that all stakeholders, even those without a background in data science, can understand our presentation. We'll focus on explaining the value and impact of our work, rather than the technical details alone.
- **Interactive Session:** Our presentation will include opportunities for interaction, including a Q&A session at the end. We'll encourage feedback and questions to ensure that all stakeholders understand and are satisfied with our work.
- **Delivery:** Our PoC and findings will be delivered to the stakeholders both in the form of this live presentation and as a comprehensive report with all our analysis, results, and recommendations. We'll also provide access to any dashboards or tools we've developed, along with necessary documentation.

Our goal with this approach is to ensure that all stakeholders understand our process, findings, and the value that our work provides to the business.

References

The anatomy of a roadmap. Product Roadmap Software. (2022, May 17).
<https://www.productplan.com/learn/roadmap-anatomy/>

Chen, J., 3DHEALS, Boelen, E., Carey, B., Likourezos, G., Kothari, F., & Genc, K. (2023, April 30). *Healthcare 3D printing, bioprinting, Innovation Platform*. 3DHeals.
<https://3dheals.com/>

Dabbura, I. (2022, September 27). *K-means clustering: Algorithm, applications, evaluation methods, and drawbacks*. Medium.
<https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a>