Predictive Marketing Analytics for Better Lead Conversion in EdTech

A Proposal report for the BDM Capstone Project

Submitted by

Name: AKHILESHWAR PANDEY

Roll number: 21F3002866

IITM Online BS Degree Program,

Indian Institute of Technology, Madras,

Chennai Tamil Nadu, India, 600036

Contents

Declaration

I hereby declare that am working on a Project Titled “Predictive Marketing Analytics for Better Lead Conversion in EdTech”. I extend my heartfelt gratitude to EdTech firm Blade Learner, for providing the necessary resources that enabled me to conduct my project.

I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered through primary sources and carefully analyzed to assure its reliability.

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through thorough analytical procedures.

I am dedicated to adhering to the information of academic honesty and integrity, and I am receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project's completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

I agree that all the recommendations are business-specific and limited to this project exclusively, and cannot be utilized for any other purpose with an IIT Madras tag. I understand that IIT Madras does not endorse this.

Signature of Candidate: Akhileshwar Pandey

Name: AKHILESHWAR PANDEY

Date: 26/01/2025

1 Executive summary

This project is based on an emerging EdTech company – Blade Learner, operating from Delhi NCR in B2C segment. It caters to the needs of humanities students primarily preparing for undergraduate entrance examinations and board exams.

Despite of having ample of operational data, Blade Learner currently lacks data science wing, and decisions they make are based on their intuitions based on limited analysis of the data. As a consequence, the firm is not getting deeper insights about their operations, sales and marketing. One of the major problems faced by the firm as they identify is in their telemarketing campaign, which they do to reach out to students and sell the courses according to needs and preferences of the students. Sales representatives try to make calls to each of the leads generated from different platforms with little insight into their likelihood of enrolling in courses resulting in Low conversion rate of student leads.

The proposed approach involves utilizing the organization's lead data to predict which students are most likely to enrol in courses and further devise efficient strategies for sales and telemarketing campaign. By applying machine learning algorithms, a predictive classification model that scores leads based on their likelihood of conversion, is to be built.

Expected outcome is to have efficient strategy for sales and telemarketing and get better lead conversion rate and hence better allocation of resources saving cost and time. Also, this would establish a foundation for a data-driven decision-making culture within the organization.

2 Organization Background

Blade Learner, established in October 2020 in the middle of COVID era by passionate educators Sudhanshu (BHU alumnus) and Ayush Raj (alumnus of Jamia and JNU), at Darbhanga, Bihar. Later on, Blade Learner shifted their operations at Delhi NCR.

Blade Learner is a dedicated platform for Humanities and Arts enthusiasts. It helps in preparing for undergraduate entrance exams and board exams. It was founded with a clear mission as stated by founders: to bridge the gap in quality educational resources for students in humanities.

With a thriving community of over 27,000 YouTube subscribers, Blade Learner offers more than thousand engaging videos, test-series, and notes tailored specifically for Humanities students. The platform has been instrumental in helping numerous students excel in competitive exams like CUET, paving the way for admissions to prestigious central universities such as

DU, JNU, BHU, Visva Bharati, and more. Its customer base is predominantly based in Hindi speaking regions in North.

3 Problem Statement

3.1 Statement 1: The organization faces challenges in identifying high-potential leads, leading to indecisive strategy for tele-calling resulting in low conversion rate despite significant outreach efforts. Without a data-driven method to prioritize leads, resources (man X hours) are inefficiently allocated.

3.2 Statement 2: Organization wants to analyse the effectiveness of different follow-up timelines and frequencies in order to get insights for further optimization of engagement strategies. Also, the organization wants to redesign the calling strategy from scratch after reviewing the sales funnel.

4 Background of the Problem

As Blade Learner is an EdTech firm curating and delivering contents for humanities students preparing for undergraduate entrance examinations, each of its operation is conducted on online platforms as a result they have a lot of data being generated from these platforms like you tube, web application, native application etc.

Customers of the Blade Learner are students of class 11^{th} and 12^{th} who have opted for humanities leaving out Science and Commerce. These students have to take permissions from their guardian to buy any course.

Blade Learner have as much data as any EdTech firm could have. For purpose pertaining to leads, the data they have comprises of different user engagement matrices such as 6-7 categories of leads, courses clicked by user on their native or web application, date of calls made, class, contact details etc. But this data is being used only for limited analysis. The calls are being made to the leads without properly gaining insights from the data collected, such as engagement metrics and inquiry types. Also, the organization is lacking efficient strategy to prioritize the telemarketing calls based on any ordering or preferences. What they do is, they try to make calls to every potential customer without any ordering. As a consequence, the conversion rate of leads remains at something less than ten percent. This results in increased operational cost and also losing customers, which is identified as bottleneck by the organization in their growth and expansion.

5 Problem Solving Approach

5.1 Proposed Methodology

Given the problem and data at disposal comprehensive quantitative and qualitative approach using proper technologies would be used.

Firstly, data would be visualised using excel, as in its raw form data has very large missing values and many irrelevant columns. Using different functions like pivot tables it's easy to get more insights. After that firsthand cleaning would be done and it would be prepared for further preprocessing steps using python.

Next step would be to perform Exploratory Data Analysis (EDA) using python. Based on this EDA featuring engineering would be done in order to prepare suitable Machine Learning Model. Further a baseline ML model would be built which would be further hyperparameter tuned to get the best ML model suited in order to predict the likelihood of conversion of leads.

Along with this, deliberation with firm on their lead pursuing strategy would be conducted on timely basis. It would be figured out that if they are following BANT(Budget-Authority-Need-Timeframe) strategy as first principle to approach the leads.

5.2 Data Collection

To address the challenges faced by Blade Learner in optimizing lead conversion rates and follow-up strategies, comprehensive Quantitative and Qualitative data collection is essential.

As the problem is related to lead conversion strategy data would be collected to know the lead details like source of lead, course clicked, demographics (age, location, class etc). This data will allow for a detailed understanding of the leads' preferences, enabling more targeted and personalized outreach strategies.

Further, Engagement metrics such as number of calls, date of first call, and follow up dates would be sought as this will help identify patterns that correlate with higher conversion likelihood, facilitating better lead prioritization.

5.3 Analysis Tools

To address the challenges and objectives outlined in this report, a combination of advanced analytical tools and techniques will be employed.

Microsoft Excel: Excel is used for first hand Preliminary examination, basic cleaning, and exploration of the dataset to navigate and understand data without writing any code. This is because excel is user-friendly, making it an ideal tool for initial data exploration and quick fixes before more advanced analysis.

Python with Libraries

Pandas and NumPy: For data preprocessing, cleaning, and transformation.

Scikit-learn: To implement machine learning tasks like splitting the data, doing feature engineering and training models like logistic regression, decision trees, and random forests for predicting conversion likelihood.

Visualization Tools (Matplotlib, Seaborn): Creating detailed visualizations for data distributions, engagement patterns etc.

6 Expected Timeline

6.1 Work Breakdown Structure

Field Survey and connecting with Businesses: Started during end of December 2024 and was conducted till 6th-7th of January.

Data Collection: Spanning about 3 days this was done until 10th of January.

Preliminary Data Processing: Preliminary processing was done till 22 January.

Proposal Preparation: Proposal was prepared right after getting the data until 26th of January.

Data Processing and Analysis (Python): would be done by the end of January until 28th.

Feature Engineering: this would be completed by 5th of February,

Model Selection, Hyperparameter Tuning, and Evaluation: expected to be completed by 5th of March.

It is planned that conclusive recommendations to the business would be provided by mid of the March 2025.

6.2 Gantt Chart

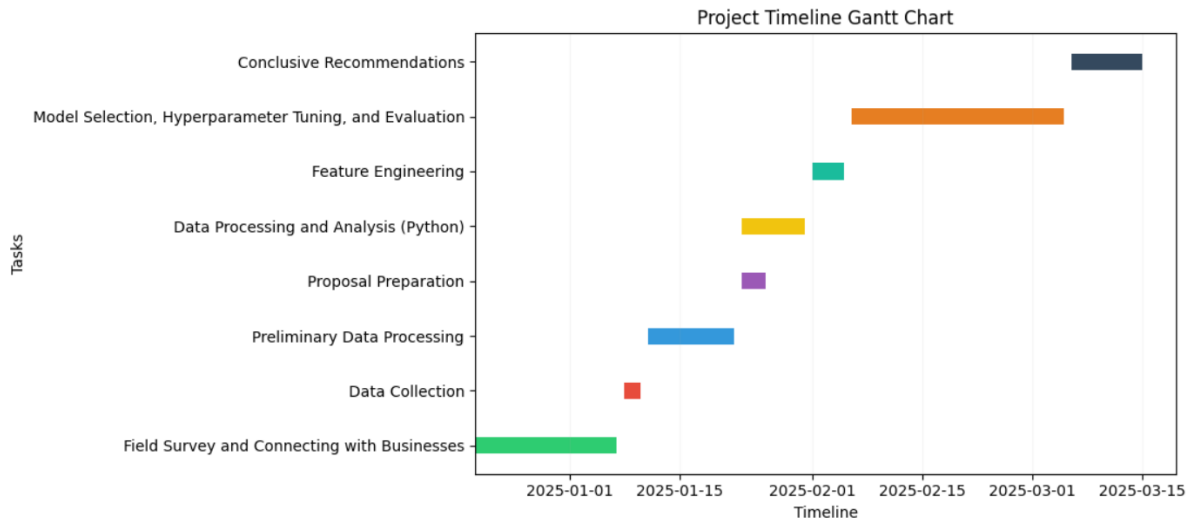

Figure 1: Expected timeline for completion of project.

7 Expected Outcome

- As a direct effect lead conversion rates should increase with prioritized tele-calls to those customers whose likelihood to convert is more as predicted by the ML model.
- To have an efficient follow up strategy in place to enable the organization to streamline its communication strategy, reducing unnecessary efforts.
- To have better future outreach strategies given insights into activities related with conversion.
- Applying tools, techniques, and methodologies implemented in this project to other organizational challenges, such as course feedback analysis, marketing ROI etc.
- To foster data-centric decision-making in order to build a foundation for analytics-driven strategies to enable the organization to approach future challenges with solutions backed by analysis of data.