

Final Group Project

Professor Ernesto Lee

Understanding the Business Context of the Final Project

The marketing head of a profit center at DTCC (where you are a data scientist) approaches you with a problem they would like to be addressed. The marketing team recently completed a marketing campaign where they have collated a lot of information on existing customers. **They require your help to identify which of these customers are likely to buy a term deposit plan.** Based on your assessment of the customer base, the marketing team will chalk out strategies for target marketing. The marketing team has provided access to historical data of past campaigns and their outcomes—that is, whether the targeted customers really bought the term deposits or not. This is your dataset that is attached (bank-full.csv in: <https://github.com/fenago/imbalanced>). Equipped with the historical data, you have set out on the task to identify the customers with the highest propensity (an inclination) to buy term deposits.

Business Discovery

The first process when embarking on a data science problem like the preceding is the business discovery process. This entails understanding various drivers influencing the business problem. Getting to know the business drivers is important as it will help in formulating hypotheses about the business problem, which can be verified during the exploratory data analysis (EDA). The verification of hypotheses will help in formulating intuitions for feature engineering, which will be critical for the veracity of the models that we build.

Let's understand this process in detail from the context of our use case. The problem statement is to identify those customers who have a propensity to buy term deposits. As you might be aware, term deposits are bank instruments where your money will be locked for a certain period, assuring higher interest rates than saving accounts or interest-bearing checking accounts. From an investment propensity perspective, term deposits are generally popular among risk-averse customers.

Equipped with the business context, let's look at some questions on business factors influencing a propensity to buy term deposits:

Would age be a factor, with more propensity shown by the elderly?

Is there any relationship between employment status and the propensity to buy term deposits?

Would the asset portfolio of a customer—that is, house, loan, or higher bank balance—influence the propensity to buy?

Will demographics such as marital status and education influence the propensity to buy term deposits? If so, how are demographics correlated to a propensity to buy?

Formulating questions on the business context is critical as this will help in arriving at various trails that we can take when we do exploratory analysis.

Finally, take the dataset and perform feature engineering.

If the data is imbalanced, then balance the data by undersampling or oversampling.

Train the data against several models and choose the model that performs the best.

Tune the hyperparameters of your model to optimize.