

CHAPTER 3 – RESEARCH METHOD

This chapter describes the approach used to analyze the variables influencing customer churn at PT XYZ and to build a successful prediction model using a Multilayer Perceptron (MLP) neural network. In this research, Figure 3.1 shows the research methodology employed by the researcher to conduct the study. The method combines extensive data analysis with powerful machine learning algorithms and marketing intelligence to create customized client retention tactics. This methodical investigation starts with a full explanation of the data gathering procedure, followed by the actions necessary to prepare the dataset for analysis. Following that, the MLP model's design and execution are described, including training, validation, and the incorporation of marketing information. The chapter lays the groundwork for using the model's predictive capabilities to influence strategic choices, while also addressing the research aims and concerns raised in Chapter 1.

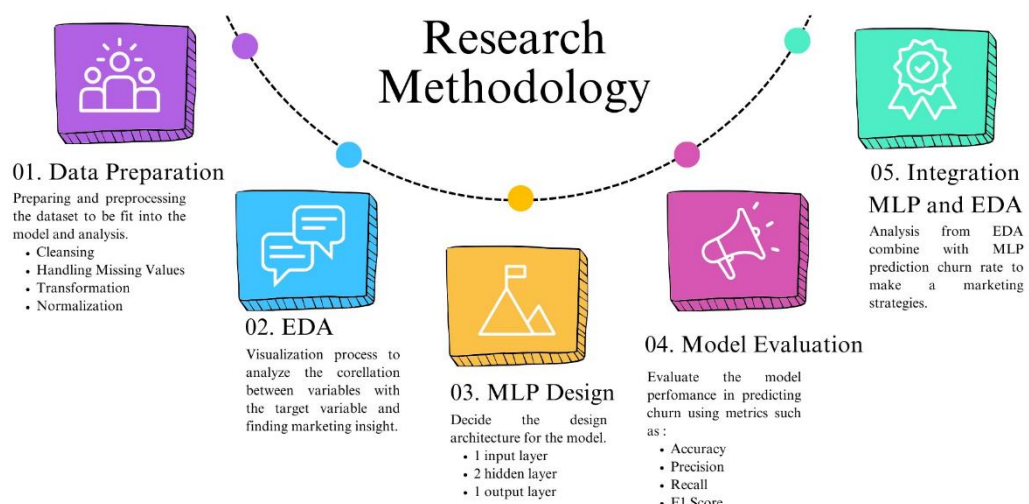


Figure 3.1 Research Methodology

3.1 Data Description

The dataset used in this research contains 84,605 records divided into 36 attributes, representing a comprehensive collection of consumer interactions and characteristics within PT XYZ. The dataset covers a period span, capturing key temporal milestones, though the precise date ranges are unknown. It's packed with demographic information like gender, geographic area down to sub-districts, and age, giving you a detailed picture of PT XYZ's consumer base.

The data types are varied, including both quantitative and qualitative features; numeric variables contain transactional information such as package costs and ticket volumes, while category categories include subscription kinds and client identities. This combination of data sets allows for a multidimensional examination of churn, which aids in the construction of a prediction model that matches the complicated reality of customer retention difficulties.

Table 3.1 PT XYZ Dataset Information

Category	Columns
Demographic Information	customergender, DOB, CSTMR_AGE
Subscription Details	DATE_ACTIVATION, DATE_EXPIRED, SUBSCRIPTION_AGE, PACKAGEPRICE_IDR, PACKAGENAME, typeRetentionDay, typeRetentionMonth
Billing and Payment Information	dayPart, monthPart, BA_number, PROFIT_CENTER_NAME,
Communication and Service Interaction	tik_id, cat1, sub1, seg, closeDate, closeTime, Total Ticket, trendcall
Geographic Information	CITY, DISTRICT, SUB_DISTRICT, Mapping_Area, SA_Province, territory
Technology and Network	Technology, profit_center
Customer Account Information	prepaidType, Customer_Name, tik_Custid, tik_CA_Number, SIA_Number
Target Variable	CHURN

3.2 Data Preparation and Preprocessing

The dataset, which served as the foundation for this investigation, was carefully preprocessed to guarantee the reliability and validity of the results. A variety of critical milestones distinguished this preliminary period. Data cleansing was initiated, methodically addressing inconsistencies and correcting mistakes to build a solid foundation. The focus then shifted to missing values, where a methodical assessment process drove judgments. Decisions were made to either impute missing data or, in certain situations, discard records or characteristics with severe gaps, always with data integrity in mind.

Outlier analysis was then performed to find data points that differed considerably from the norm. Each outlier was examined in its own context, and judgments were made on whether to alter or keep it depending on its contribution to the research's understanding. The significance of data consistency was understood, and as a result, data transformation was commenced. To achieve consistency, numerical data were standardized, which is critical for the performance of the neural network model. Encoding methods were used to transform categorical information into numerical representation to simplify computer analysis.

The dataset was finally divided into separate training and testing sets using the random split technique. This important step allows for accurate model assessment and protects against overfitting. A strong basis for the challenging analytical stages of the investigation has been built through these cautious and thoughtful preparations.

3.3 MLP Neural Network Design

The Multilayer Perceptron (MLP) neural network, a critical component of the research, was carefully designed to match the specific features of the dataset and the difficulty of the customer churn prediction problem. A data-driven method was used to design the architecture, beginning with a simple arrangement and prioritizing simplicity over complication. A single hidden layer was suggested for datasets with reasonably simple patterns. However, for more complex issues involving the nuanced prediction of customer churn, the need for further depth was observed. As a result, one or two more

hidden layers were added, allowing the network to capture and simulate complex interactions.

A critical hyperparameter, the number of nodes inside each layer, was carefully evaluated. A small number of nodes were initially used to establish a compromise between model expressiveness and efficiency. For example, with 36 input features in the dataset, early hidden layers included 36 to 72 nodes. This decision enabled the network to accommodate the complexity of the data without adding excessive complexity.

The activation functions, which drive neural network non-linearity, were carefully designed. Rectified Linear Unit (ReLU) activation functions were used to enhance the network's capacity to understand complicated patterns and correlations within the data. The output layer was chosen based on the nature of the issue. The sigmoid activation function was used in the situation since binary classification (churn vs. no churn) was the aim.

The neural network architecture is the foundation for predictive modeling, and its design represents a careful balance between data-driven flexibility and neural network theoretical principles. Experimentation and hyperparameter tweaking will continue to develop this approach, ensuring that it corresponds with the complicated patterns inherent in the customer churn prediction challenge.

3.4 Exploratory Data Analysis (EDA)

In this phase begin with an in-depth trip through the dataset in this step of Exploratory Data Analysis (EDA) to discover vital marketing insights. The process starts with the use of histograms and bar charts, which enable us to graphically explore the distribution of categorical data such as customer gender, churn status, and service bundles. These visualizations give an instant, intuitive understanding of the underlying trends and inequalities within these categories.

Further investigation employs scatter plots and box plots for a more in-depth look at continuous data, offering insight on patterns, outliers, and probable anomalies that may influence consumer behavior. This is critical in determining elements that may impact

churn or customer satisfaction. Heatmaps are used to highlight the strength and type of relationships between various variables. This graphic tool is especially good at illustrating linkages that aren't immediately obvious. Time-series analysis is used to investigate how specific tendencies change over time, revealing information about customer lifetime and subscription habits.

This comprehensive EDA strategy is about translating data into usable marketing insights, not merely knowing the facts. Customized approaches that not only minimize churn but also increase customer happiness and loyalty can be built by methodically evaluating consumer behavior, preferences, and engagement. As a result, this phase serves as a vital link between raw data and strategic marketing decision-making, ensuring that every action is influenced by solid, data-driven insights.

3.5 Model Evaluation

In the Model Evaluation section, the performance of the MLP model will be thoroughly examined using multiple metrics and validation approaches. The evaluation process starts with the establishment of essential metrics such as accuracy, precision, recall, and the F1 score. The use of a confusion matrix will enhance this research by offering a thorough perspective of the model's ability to predict each class properly. This matrix will not only supplement the overall accuracy evaluation, but it will also provide deeper insights into the balance of precision and recall, improving comprehension of the model's ability to recognize genuine positives and the usefulness of the F1 score in balancing these elements.

Validation approaches are crucial in guaranteeing the model's adaptability and generalizability. To evaluate the model's performance across multiple subsets of the dataset, cross-validation, specifically k-fold cross-validation, will be used. This strategy reduces the danger of overfitting while also providing a more trustworthy indicator of how the model will perform on untested data. Other validation approaches, such as a hold-out validation set or bootstrapping, may also be explored to increase the resilience and dependability of the model.

Overall, this comprehensive assessment approach is intended to not only quantify the MLP model's efficacy, but also to fine-tune its parameters for optimum performance, ensuring that the predictions it provides are both accurate and dependable in a real-world setting.

3.6 Integration of MLP and EDA Findings

The purpose of the Integration of MLP and EDA Findings section is to combine the insights gained from EDA with the predictive capabilities of the MLP model. The EDA gives a comprehensive picture of consumer behavior and preferences, identifying patterns and trends critical to focused marketing efforts. These findings will supplement the MLP model's churn forecasts, providing a more detailed picture of consumer behavior.

Using EDA results to enhance and concentrate the MLP model's prediction capabilities is one strategy for merging these insights. Customer segments discovered by EDA, for example, may be utilized to tailor the model's training, making its predictions more relevant to certain groups. Furthermore, EDA may aid in the interpretation of the MLP model's outputs by giving context for the predicted patterns it reveals.

In the end, this integrated strategy improves the capacity to develop successful client retention tactics. More focused, data-driven marketing interventions may be generated by integrating predictive analytics from the MLP model with deep market insights from EDA, resulting in enhanced client retention and satisfaction.