# Title

Predicting Customer Churn in DTH company with Machine Learning:

A Novel Approach to Enhancing Customer Retention Strategies

**Executive Summary**

This research focuses on addressing the critical issue of customer churn in the highly competitive DTH sector . With the growing challenges in retaining customers, it has become imperative for companies to leverage advanced predictive analytics to anticipate churn risks. This study carves out a niche by developing a machine learning-based model for predicting customer churn, utilizing a rich dataset comprising various customer interactions, transactions, and behavioural patterns. The aim is to not only predict potential churners with high accuracy but also to provide actionable insights that can be integrated into strategic decision-making processes for improving customer retention.

**Objectives and Scope**

1. **To develop a predictive model** that accurately identifies customers at high risk of churn, enabling the company to implement targeted retention strategies.
2. **To analyse customer behaviour and interactions** to uncover underlying patterns associated with churn, thereby facilitating more informed business decisions and personalized customer experiences.
3. **To differentiate the proposed solution** from existing models by incorporating a broader set of predictive variables and employing advanced machine learning techniques, resulting in improved prediction accuracy and actionable insights for retention strategies.

The significance of this project lies in its potential to transform how DTH companies approach customer retention, shifting from reactive measures to proactive, data-driven strategies.

**Methodology**

The methodology encompasses several key phases: data collection, pre-processing, exploratory data analysis (EDA), model development, evaluation, and implementation of findings into business strategies.

1. **Data Loading and Pre-processing**: Load the new dataset, handle missing values if any, encode categorical variables, and normalize numerical features as necessary.
2. **Exploratory Data Analysis (EDA)**: Understand the distribution of variables, especially the target variable 'Churn', to gain insights into the dataset.

Methods involve statistical analysis and visualization to understand the distribution of variables, identify correlations, and detect patterns or anomalies in the data.

1. **Model Training**: Select an appropriate model for classification (e.g., Random Forest, Gradient Boosting) and train it on the dataset.
2. **Model Evaluation**: Evaluate the model using appropriate metrics to ensure its performance.
3. **Feature Importance Analysis**: Determine which features are most predictive of churn to inform our campaign recommendations.
4. **Business Recommendations**: Based on the model's insights, develop targeted strategies for retention campaigns that are cost-effective and designed to pass the revenue assurance team's scrutiny.

**Type of Research**: This is an Empirical research project, focusing on the analysis of data to uncover patterns and test hypotheses about customer churn.

**Data Collection Methods**: The study utilizes Secondary data collected from the company's internal databases, including customer demographics, transaction histories, interaction logs, satisfaction scores, and churn records.

* **Data Pre-processing**: Techniques include handling missing values through imputation, encoding categorical variables, normalizing numerical features, and removing outliers to prepare a clean dataset for analysis.
* **Exploratory Data Analysis (EDA)**: Methods involve statistical analysis and visualization to understand the distribution of variables, identify correlations, and detect patterns or anomalies in the data.
  + **Machine Learning Algorithms**:
  + Supervised Learning: Random Forest and Gradient Boosting classifiers for churn prediction, due to their ability to handle the complex relationships between variables.
  + Unsupervised Learning: K-Means clustering for segmenting customers into distinct groups based on their behaviour and characteristics, which can inform targeted retention strategies.

The research integrates these methodologies to develop a comprehensive approach to predicting customer churn and supporting the development of effective retention campaigns.

**Preliminary Findings & Expected Results**

Existing models in the domain of customer churn prediction primarily utilize classic machine learning algorithms, such as logistic regression, decision trees, and simple ensemble methods. These models often focus on a limited set of features, such as tenure, billing information, and basic demographic data, to predict churn.

The proposed research is expected to advance beyond these models by incorporating a wider array of predictive features, including detailed interaction data, satisfaction scores, and behavioural patterns. By leveraging advanced machine learning techniques, such as Gradient Boosting and Random Forest, coupled with sophisticated data pre-processing methods, the expectation is to achieve higher accuracy in churn prediction. Additionally, the utilization of unsupervised learning for customer segmentation is anticipated to provide deeper insights into customer behaviour, further enhancing the effectiveness of targeted retention strategies. The ultimate goal is to offer a novel solution that not only predicts churn with high accuracy but also delivers actionable insights for crafting more effective, data-driven customer retention strategies.