MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

Model optimization of Customer Churn and Automation of Cross Sell Impact Pipeline

A Report On Industrial Training

At

UPL Ltd. Bangalore

Date: 05/06/2023-05/08/2023

Submitted by

Vanshika Gupta

200968118

CERTIFICATE

UPL Limited, UPL House,
610 B/2, Bandra Village
Off Western Express Highway
Bandra (East), Mumbai 400 051, India.

w: www.upl-ltd.com
t: +91 22 7152 8000

28th August '2023

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Ms. Vanshika Gupta has successfully completed her internship at UPL Limited from 5th June '2023 to 4th August '2023 under the NextGen Internship Program. Her project location was Bangalore and she worked in the Digital & Analytics.

During this period she has worked on the project "Code Optimization" under the guidance of Aakshi Sharma. Her deliverables involved:

- Streamlining the Cross Sell Mexico Code for automation
- Built wireframe for recommendation dashboard

As an intern she has completed the assignments, within the stipulated timelines and demonstrated professionalism, dedication, and worked in a collaborative manner.

We were pleased to have Vanshika Gupta as a part of our team to create her summer story and we wish her great success for future endeavours.

Yours sincerely,
For **UPL Limited**

Santosh Vellanki
Head HR - Corporate

DECLARATION

I hereby declare that this Industrial Training project work entitled Customer Churn and Cross sell automation is original and has been carried out by me at UPL Ltd., Bangalore, under the guidance of Aakshi Sharma,Senior data scientist at UPL. No part of this work has been submitted for the award of a degree or diploma either to this University or to any other Universities.

Your Signature: 

Your Name: Vanshika Gupta

Reg. No: 200968118

Place: MIT, Library

Date : 05/08/2023

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to UPL Ltd for providing me with the opportunity to work as a summer intern on the project "Customer Churn and Cross Sell Automation." These two months of internship have been a valuable learning experience, allowing me to apply my academic knowledge in a real-world setting.

I am deeply thankful to my mentor, Aakshi Sharma, whose guidance, and support were instrumental throughout this project. Aakshi's expertise and encouragement motivated me to push my boundaries and achieve the project goals. Her insights and feedback have been invaluable in shaping my understanding of customer churn and cross sell strategies.

I am also thankful to the entire team at UPL Ltd for creating a positive and conducive work environment. The collaborative spirit of the team enhanced my learning experience, and I am grateful for the opportunities to collaborate, learn, and grow.

Lastly, I would like to express my appreciation to everyone who contributed to my internship journey, both directly and indirectly. Your support and encouragement have been crucial in making this internship a fulfilling and enriching experience.

ABSTRACT

In the customer churn project, we addressed the difficulty of forecasting customer attrition in the agriculture industry, with a focus on identifying hard and soft churn for UPL, a prominent agricultural enterprise. With more than 230 features, the dataset presented a serious dimensionality issue. Our aims included feature selection and dimensionality reduction with methods such as PCA; accuracy enhancement for hard and soft churn via machine learning algorithms; interpretability and domain relevance with assistance from agricultural specialists; computational efficiency and scalability optimisation of the model; and a structure for ongoing iteration and enhancement. By achieving these goals, our study sought to develop an accurate, comprehensible, and flexible predictive model suited to the particular dynamics of consumer behaviour in the agriculture sector, offering insightful information for strategic decision-making and customer retention efforts at UPL.

The goal of the automation of cross sell project was to automate a corporation that deals with agricultural by means of a cross-sell pipeline. Its goal was to replace the manual lead matching process with an automated system that was data-driven and sophisticated. The main goals of the project were to establish a user-friendly dashboard for real-time insights, efficient data processing and preprocessing, and innovative algorithms for lead matching. To improve customer happiness, the project also gave personalization top priority when making cross-sell recommendations. The initiative aimed to enhance operational efficiency, boost conversion rates, and stimulate revenue growth via successful cross-selling by accomplishing these objectives. A strong, flexible, and customer-focused strategy to cross-selling was ensured by data security, compliance, and a dedication to ongoing improvement.

CONTENTS

List of tables

List Of Figures

Introduction

1.1 Introduction about first project-Model Optimisation for Customer Churn

1.1.1. Section 1

In the business world, the term "customer churn" describes the occurrence where customers cease purchasing or utilizing a company's goods or services. For businesses, this can be somewhat concerning because it's usually less expensive to keep current clients than to find new ones. Hard and soft churn are the two basic categories into which customer attrition falls.
Hard Churn: When clients entirely quit utilising a business's goods or services, this is known as hard churn. Hard churn in the context of UPL, an agricultural corporation, would be farmers or agricultural enterprises giving up on using UPL's seeds, fertilisers, and insecticides.
Conversely, soft churn describes circumstances in which clients spend less money or engage with the business less frequently yet continue to use its goods and services. Customers that reduce their purchases of agricultural products or look into alternatives while still making occasional purchases from UPL may be said to be experiencing mild churn.
Significance to the Current Global Situation-
For agricultural organizations like UPL, understanding customer turnover is critical in the current world for a number of reasons:

Sustainability and Environmental Concerns: Farmers may choose eco-friendly or organic alternatives as a result of growing awareness of sustainable agriculture methods and environmental concerns. This could cause soft churn for businesses who make agricultural products derived from chemicals.

Technological Developments: The agriculture sector is seeing several technological developments, such as AI-driven solutions, IoT devices, and precision farming. When farmers use these technologies, their preferences may shift, which could have an impact on how they interact with conventional agricultural product suppliers like UPL.

Climate Change: Agricultural practices are impacted by climate change. The crops and products that farmers choose to use can be impacted by changes in weather patterns, droughts, or floods.

Market Competition: As local and foreign firms become more competitive in the agricultural sector, consumers may explore alternative possibilities, which could cause either gentle or hard churn.

Regulatory Shifts: Modifications to laws governing pesticide use, environmental standards, or agricultural practices may force farmers to reevaluate the products they offer, which may result in churn.

1.1.2. Shortcomings and Motivation

Shortcomings: The curse of dimensionality presents a significant obstacle when managing a large number of features in a machine learning model. There is a chance of overfitting since the data space gets increasingly sparse as the number of features rises. Poor performance on unseen data results from overfitting, which happens when the model learns to memorise the training data instead of making inferences from it. Furthermore, as the number of characteristics increases, the model's complexity climbs as well, making it challenging to understand the judgements the model makes. This lack of interpretability is a major flaw, particularly in applications where it's important to comprehend the logic behind the predictions. The procedure is resource-intensive since training models with a large feature set requires a significant amount of time and computer power. Lastly, the presence of irrelevant or redundant features further hampers the model's accuracy, as it may focus on noise rather than meaningful patterns in the data.

Motivation: Machine learning models are highly motivated to be improved by the difficulties presented by a plethora of features. It is crucial to use feature selection and engineering techniques in order to combat the problems caused by dimensionality and overfitting. The accuracy and generalizability of the model are improved by determining the subset of features that have a substantial impact on the predictions. High-dimensional data can be lessened in impact while maintaining important information by using dimensionality reduction techniques like PCA. Furthermore, the inability to be interpreted clearly encourages the investigation of models intended to be more transparent, guaranteeing that interested parties can understand and rely on the model's conclusions. Working together with domain specialists is essential since their expertise can help direct the selection of features and improve the model's applicability to the particular industry, like agriculture. Furthermore, it becomes clear that machine learning projects are iterative as overcoming the difficulties presented by a large feature set requires constant testing, optimisation, and refinement. The goal of these initiatives is to develop models that overcome the inadequacies and provide insightful information for decision-making. These models should be accurate, interpretable, relevant, and computationally efficient.

1.1.3. Problem Definition & Objectives

Problem definition:

Predicting client attrition in the agriculture industry—more especially, distinguishing between hard and soft churn for UPL—is the central challenge of this study. Hard churn is when a consumer completely stops doing business with UPL, whereas soft churn is when a customer reduces their interaction with UPL, maybe looking into other options but not completely stopping using UPL's goods and services. The dataset has an astounding 230+ features, which presents a serious dimensionality issue. The process of modelling is made more difficult by the number of characteristics, which also increases the risk of overfitting—the model learning the training set and not adapting well to new, unobserved data. Furthermore, the model must be customised to the particular dynamics of the agriculture sector in order to guarantee its applicability and precision in forecasting consumer behaviour.

Project Objectives:

- **Feature Selection and Dimensionality Reduction:** Reducing the large feature collection to a manageable, informative subset is the main goal. The objective is to determine the essential features required for reliably predicting both hard and soft churn through the use of methods such as Principal Component Analysis (PCA) and feature selection algorithms. Reducing the number of dimensions in the model allows it to concentrate on the most important elements of consumer behaviour, improving forecast accuracy.
- **Accuracy Improvement for Hard and Soft Churn: The** goal of the project is to make predictions for both hard and soft churns more accurate. To determine which machine learning method is most appropriate for this particular issue, a number of them will be tried and refined. Cross-validation and hyperparameter tweaking are two strategies that will be used to make sure the model is reliable and robust in capturing the intricacies of customer turnover in the agricultural industry.
- **Interpretability and Domain Relevance:** It's critical that the predictions made by the model can be understood, particularly by stakeholders and subject matter experts. Working together with agricultural experts is essential to guaranteeing that the characteristics chosen fit the specifics of the sector. Enabling strategic decision-making for customer retention initiatives, an interpretable model not only makes accurate predictions but also provides insights into the reasons behind the churn of specific customers.
- **Optimisation and Scalability:** The model must be optimised for computational efficiency in addition to accuracy and interpretability. Large-scale agricultural datasets should be handled by its scalability, enabling real-time or nearly real-time prediction making. The applicability of the model in

realistic, dynamic circumstances is ensured by optimising its computational complexity.

- **Iteration and Improvement Constant:** The initiative recognises that consumer behaviour is dynamic and that the agricultural business is changing. Thus, creating an iterative process is one of the main goals. This entails keeping an eye on the model's performance continuously, adding new data, and modifying the model to account for shifting consumer patterns. The model's accuracy and applicability are maintained throughout time via regular updates and improvements.

By tackling these goals, the project hopes to develop a predictive model for customer turnover in the agriculture industry that is accurate as well as interpretable, pertinent, and flexible. This all-encompassing strategy guarantees that the model not only offers precise forecasts but also insightful information for strategic decision-making in UPL's client retention initiatives for both soft and hard churn situations.

1.2. Introduction about second project- Automation Of Cross Sell Impact Calculation Pipeline

1.2.1. Section 1

The goal of this project was to transform UPL's cross-selling strategies. There were many difficulties in the manual process of matching supplied leads with converted clients, which could have resulted in errors and inefficiencies. As a result, the project's goal was to fully automate this pipeline, from loading data to preprocessing and conversion rate calculation. Enhancing the accuracy and effectiveness of UPL's cross-sell strategies was the goal of the project, which involved replacing this manual method with an automated and simplified system. The development of an extensive dashboard that offered real-time information and analysis was essential to this endeavour.

1.2.2. Shortcomings and motivations

Shortcomings: The cross-sell pipeline automation process at UPL has a few possible drawbacks that should be taken into account. First off, issues with the quality of the data may arise for the project. Erroneous recommendations resulting from missing or inaccurate data may affect the overall efficacy of cross-selling tactics. Furthermore, the degree of customisation that automation may achieve may be limited. Automated systems may find it difficult to comprehend the subtle tastes of particular clients, which could lead to generic recommendations that are poorly received. Moreover, there can be restrictions on the algorithms used for cross-selling and lead matching. Insufficient sophistication of these algorithms

may cause them to fail in recognising complex client behaviours, which would impair the accuracy of the suggestions.

Motivations: There are several different reasons for automating the cross-selling processes. Among these, the goal of time and efficiency savings comes first. The organisation may greatly increase its operational efficiency by automating the manual lead matching procedures. This will free up time and resources that can be allocated to strategic initiatives. Furthermore, automation promises to increase cross-sell process accuracy. With the help of algorithms and data-driven insights, the system can reduce the possibility of human error, guarantee accurate lead matching, and raise the possibility of profitable cross-sell conversions. Another important incentive is an improved customer experience. Cross-selling advice that are timely and tailored to each customer's demands will increase customer satisfaction and loyalty. Automation also makes data-driven decision-making easier. In the end, automation facilitates effective cross-selling, which is a driver of revenue growth that helps the business increase overall sales figures and optimise the value of its current customer base.

1.2.3. Problem Definition and Objectives

Problem Definition:

The internship project sought to automate and simplify UPL's current cross-selling funnel in order to meet this problem. The project took a complete approach, involving preparing the data, developing and importing functions, loading data, and figuring out the lead conversion rates that the team had supplied. The principal aim was to fully automate the process, therefore removing the necessity for manual involvement and guaranteeing precise and prompt matching of leads with prospective cross-sell opportunities.

In this sense, "leads" refers to the team's recommendations that highlight clients who are likely to leave. Offering comparable or better items to these clients in an effort to keep them as loyal customers and keep them from moving to competitors was the cross-sell strategy. The project's goal was to automate this procedure in order to improve customer satisfaction by swiftly offering customised product recommendations while also boosting UPL's cross-sell operations' efficiency.

One important result of this effort was the development of an easy-to-use dashboard. The cross-sell operations were tracked and analysed via this dashboard, which acted as a central location. The intern aimed to enhance UPL's cross-selling efforts with this project, guaranteeing a smooth and data-driven strategy for client retention and income production. The report's next sections will explore the approaches used, problems encountered, solutions created, and overall effects of the automated cross-sell pipeline on UPL's customer involvement and business operations.

Project objectives:
Optimising and streamlining the current lead matching process was the main goal of UPL's cross-sell automation project. This required switching from an entirely manual process to one that was fully automated. The particular goals comprised:

- **Automation of Lead Matching:** Create algorithms and protocols to automate the process of matching prospective clients with company-provided leads, guaranteeing precise cross-selling opportunities identification.

- **Data Processing and preparing:** Use methods for loading, cleaning, and preparing data that are effective. Accurate and well-structured data is necessary for accurate lead matching and trustworthy cross-selling suggestions.

- **Algorithm Development: Create** and apply cutting-edge algorithms to examine client information, online activity, and past purchasing trends. These algorithms would make it easier to create customised cross-sell recommendations based on the unique profiles of each customer.

- **Conversion Rate Analysis:** Provide systems for computing and analysing conversion rates in order to assess how well cross-selling tactics are working. The purpose of this analysis was to evaluate how well the automated system converted leads into sales.

- Da**shboard Creation:** Create a clear, user-friendly dashboard that allows you to see cross-selling information in real time. The dashboard functioned as a central location for tracking lead matching activities, offering insights into consumer interaction, and assessing the cross-sell campaigns' overall effectiveness.

- **Personalization and Customer Satisfaction:** To increase customer satisfaction, concentrate on making cross-sell recommendations that are more personalised. Customising product recommendations according to personal preferences was done in an effort to boost consumer loyalty and engagement.

- Streamline the cross-selling procedure to improve operational efficiency in order to increase efficiency. The project's goal was to save time and money by automating repetitive operations and minimising manual intervention, freeing up UPL to concentrate on important business objectives.

- **Data Security and Compliance:** To guarantee the integrity and confidentiality of consumer data, put strong data security measures in place. Upholding industry norms and pertinent data protection laws was essential to preserving client confidence and legal compliance.

- **Continuous Improvement:** Provide a structure for ongoing observation and development. To ensure that cross-sell algorithms and strategies remain effective over time, regularly analyse performance indicators, customer feedback, and market developments.

- **Knowledge Transfer:** Provide thorough documentation of the entire procedure, including in-depth explanations of the automated cross-selling mechanism. Enable UPL's teams to comprehend, maintain, and improve the automated solution by facilitating knowledge transfer inside the company.

Background Theory / Literature Review / Related Work

During my internship at UPL, I thoroughly investigated cutting-edge machine learning (ML) and deep learning (DL) models in addition to fundamental data science methods. Using complex algorithms like Random Forest, LightGBM Classifier, CatBoost, and XGBoost was part of my job. Using methods like GridSearchCV and RandomizedSearchCV, I carefully adjusted the hyperparameters of these models to make sure they were set up for our particular jobs. In order to obtain understanding of the relevance of distinct features inside the models, I also made use of the capabilities of SHAP (SHapley Additive exPlanations) feature importance analysis. With the help of this multidimensional strategy, I was able to efficiently automate the cross-sell pipeline and give UPL a data-driven, lead matching and cross-selling strategy solution. I was able to improve UPL's cross-sell operations by applying state-of-the-art approaches and refining my model selection, optimisation, and interpretation skills through these endeavours.

Models for Machine Learning:

- **Random Forest:** During training, numerous decision trees are constructed using this ensemble learning technique, which yields a class that is the mean of the classes of each individual tree. By taking the average of several decision trees' predictions, it increases precision and decreases overfitting.

- **LightGBM Classifier:** Using tree-based learning techniques, LightGBM is a gradient boosting framework. It scales well to huge datasets and is efficient in its architecture. LightGBM divides the tree leaf-wise as opposed to level-wise, which results in quicker training times and increased effectiveness.

- Another gradient boosting library that performs exceptionally well at supporting categorical features is called **CatBoost.** It does categorical variable handling automatically, removing the need for human preprocessing. CatBoost is a powerful option for a variety of datasets since it adapts to the complexity of the data.

- **XGBoost:** A popular gradient boosting library used in machine learning contests, XGBoost is exceptionally efficient. It is well-known for its speed and performance, implementing regularised boosting, which makes it a popular option in predictive modelling.

Methods for Hyperparameter Tuning:

- **GridSearchCV**: GridSearchCV evaluates every conceivable combination of hyperparameter values while conducting a thorough search over a given parameter grid. By comparing performance indicators, it assists in determining the optimal set of hyperparameters for a model.

- **RandomizedSearchCV:** This tool conducts a randomised search across a given hyperparameter space. It assesses a collection of hyperparameters that it randomly selected from the given distributions. This method is useful in large search spaces because it effectively investigates a wide range of combinations of hyperparameters.

Methods of Feature Importance:

Shapley Additive exPlanations, or SHAP relevance of Features: A consistent way to quantify feature relevance is by SHARP values. They can be used to interpret how various features affect the model's predictions and to explain the output of any machine learning model. A thorough knowledge of the contributions of features to model outcomes is provided by SHAP values, which provide a game-theoretic approach to feature importance.

Combining these models and methods with feature importance analysis and sophisticated hyperparameter tweaking allows for a reliable and data-driven way to automating the cross-sell pipeline. Within UPL's operational framework, the selection and optimisation of these models are critical to guaranteeing precise lead matching and successful cross-selling tactics.

Design/Methodology

1. Model optimization for customer churn

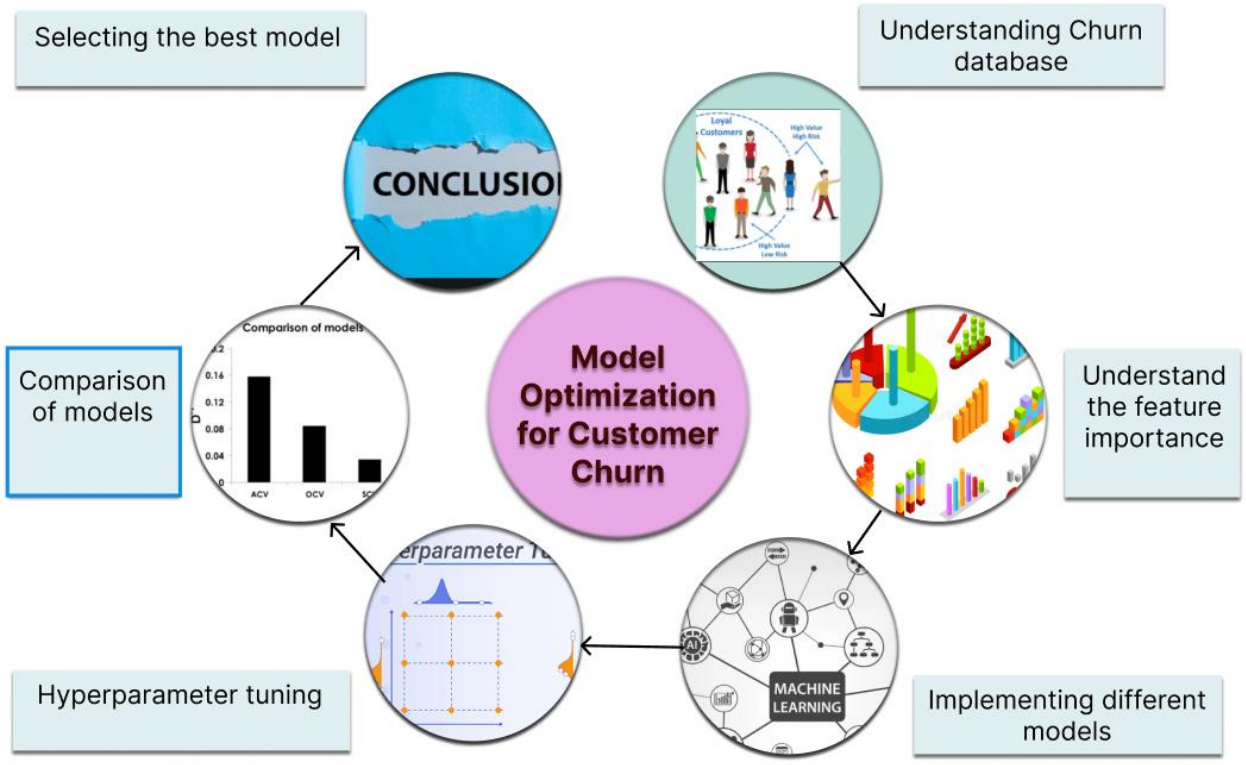

Figure-3.1 Flow diagram of Model Optimization For Customer Churn

Explanation-

We begin by looking over the churn database, which is made up of more than fifty excel pages. This contains data categories, each of which has five to ten subcategories, including the following:

1. Transactional primary data

2. Data on macroeconomics

3. Mappers to crop distribution

4. Mapping products to crops

5. Climate data

Understanding the features is the next step, which includes-

1. The significance of every feature for the model's accuracy

2. Association among characteristics

3. Classifying features according to their significance

4. Which features can be eliminated to reduce the number of features from more than 250 to more than 80.

5. Is it necessary to create new features based on pre-existing features?

We carried out feature engineering based on all of these factors. As a result, we narrowed down the feature set to about 90 features that were crucial for model prediction.

We moved on to model building once the feature-4 engineering portion was finished. We experimented with other models, such as:

1. Random forest

2. LightGBM Classifier

3. Adaboost

4. Catboost

5. XGBoost

6. Logistic regression

Different models were created for Hardchurn and Softchurn. Different datasets were fed to each of them because it was discovered that several features crucial to the hardchurn model were irrelevant to the softchurn model. After that, we used GridSearchCV and RandomizedSearchCV separately to apply hyperparameter tuning. GridSearchCV outperformed RandomizedSearchCV in its results.

We concluded that Random Forest, LightGBMClassifier, XGBoost, and Catboost were providing about equal accuracy after training and testing every model. Therefore, any of these can be applied to provide accurate forecasts.

2. Automation of cross sell impact calculation

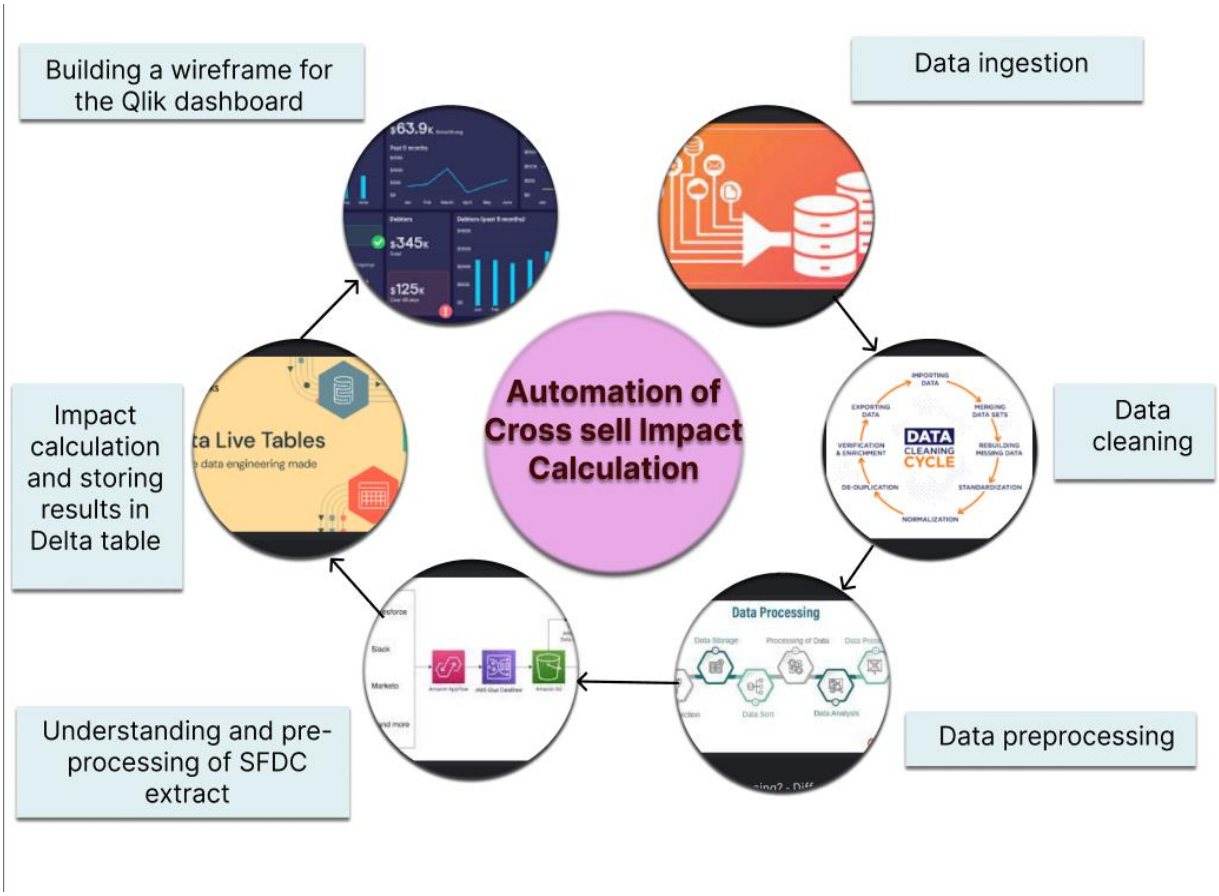

Figure-3.2 Flow diagram of Automation Of Cross Sell Impact Calculation Pipeline

Explanation-

We already have functions and pipelines in place for calculating the impact of the impact model, but the issue we were dealing with was that every time a lead was shared by the data science team at UPL, it was necessary to manually verify whether or not the lead had been converted by cross-referencing the information from delta tables with the SFDC extract of the converted leads. This was required in order to automate the impact calculation process, which further indicates how much impact is resulting from the leads that the team has shared.

My job was to automate every step of the pipeline, from data intake to impact computation and dashboard display of the final findings by connecting it to the delta tables.In order to complete this procedure, a single pipeline was created, in which data from the SFDC extract was ingested after it had been cleaned and preprocessed. comparing the leads later on and figuring out the impact from them. In order to enable easy analysis and comprehension of the current Impact scenario by the concerned teams and the target customers, the findings have finally been stored in delta tables and connected to the dashboard.

3. Impact Analysis Dashboard

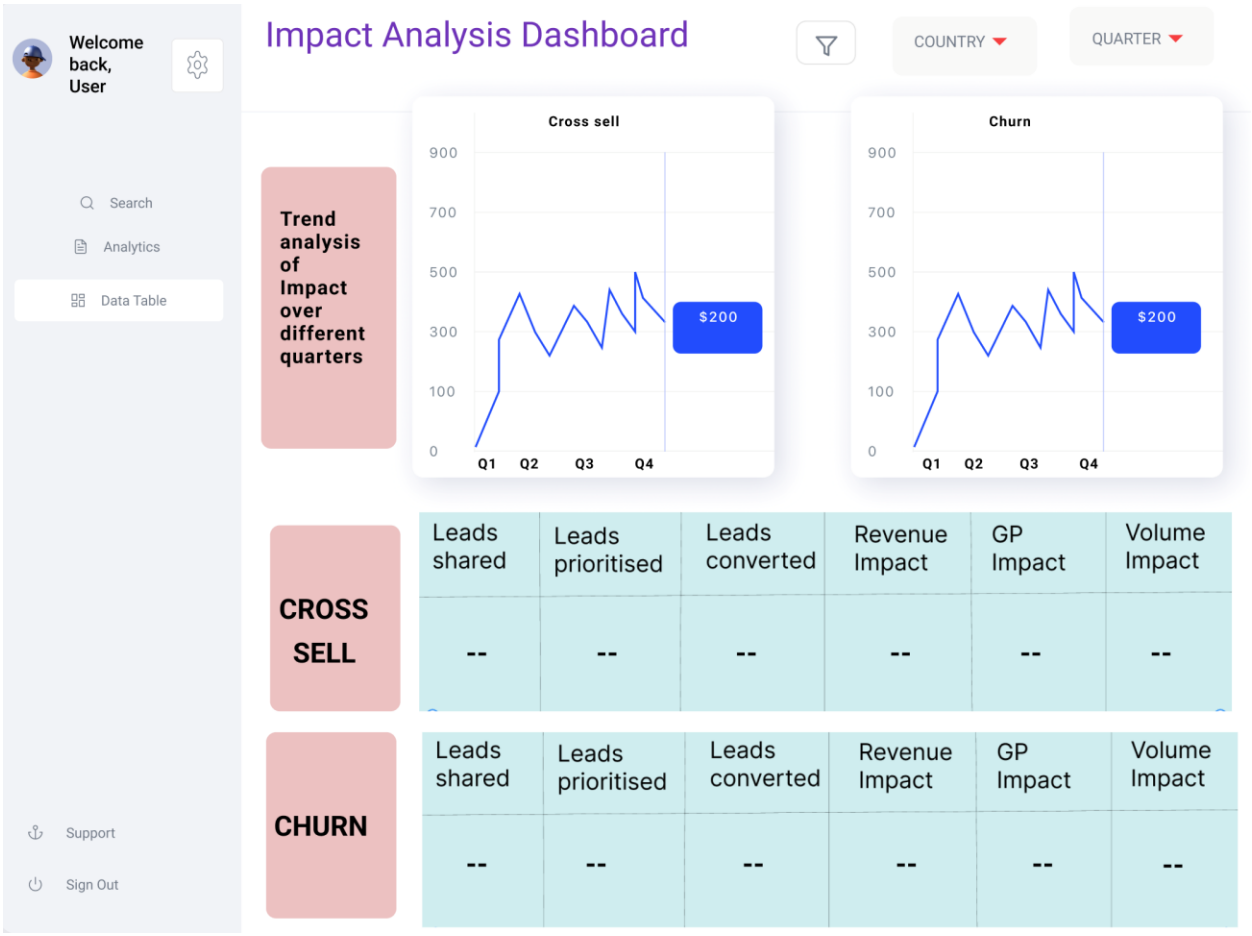

Figure-3.3 Wireframe for Dashboard -Displaying the results of the Impact Calculation

Explanation-

This dashboard gives a thorough overview of our cross-sell initiatives by visually representing the final outcomes that are kept in the delta table. Real-time updates are guaranteed by the interaction with the delta table; the dashboard automatically refreshes and reflects any modifications made to the delta tables. Our team and stakeholders will always have access to the most accurate and up-to-date insights thanks to this dynamic connectivity between the user interface and the analytical backend. The dashboard greatly aids in UPL's well-informed decision-making processes while also improving the project's outputs' accessibility. By taking a comprehensive approach, the automation of our cross-sell pipeline strengthened UPL's capacity to make data-driven, strategic decisions in real-time while also optimising operational efficiency.

Testing/Results

Soft churn model

1. Random forest model

Hyperparameters	Model Accuracy
max_depth=10, max_features=7, min_samples_leaf=20, min_samples_split=20, n_estimators=20	0.7693126022913257
bootstrap=False, max_depth=12,max_features=12, min_samples_leaf=10, min_samples_split=10,n_estimators=40	0.8062102506546951
max_depth=9, max_features=9, min_samples_leaf=20, min_samples_split=30, n_estimators=30	0.7880523731587561
max_depth=11, max_features=10, min_samples_leaf=20, min_samples_split=20, n_estimators=70	0.8229517396184063
bootstrap=False, max_depth=10, max_features=7, min_samples_leaf=20, min_samples_split=50,n_estimators=20	0.7960624766180322

Table- 4.1 Soft churn Random Forest classifier results

2. LightGBM Classifier

Hyperparameters	Model accuracy
learning_rate=0.003, max_depth=10, metric='binary_logloss',min_data=50, num_leaves=10, objective='binary', sub_feature=0.5	0.8354844743733634
learning_rate=0.02, max_depth=10, metric='binary_logloss',min_data=10, num_leaves=40, objective='binary', sub_feature=0.5	0.8192106247661803
learning_rate=0.001, max_depth=10, metric='binary_logloss',min_data=10, n_estimators=50, num_leaves=10,	0.7782694198623402

objective='binary',sub_feature=0.5	

Table-4.2 Soft churn LightGBM classifier results

Hard churn model

1. Random Forest

Hyperparameters	Model accuracy
max_depth=3, max_features=5, min_samples_leaf=70,min_samples_split=70, n_estimators=30	0.7697694776772688
max_depth=4, max_features=4, min_samples_leaf=30,min_samples_split=50, n_estimators=30	0.7812955938138314
max_depth=3, max_features=2, min_samples_leaf=50,min_samples_split=50, n_estimators=20	0.779009823947087

Table-4.3 Hard churn Random Forest classifier results

2. LightGBM Classifier

Hyperparameters	Model accuracy
learning_rate=0.003, max_depth=4, metric='binary_logloss',min_data=50, n_estimators=50, num_leaves=10, objective='binary',sub_feature=0.5	0.792308532721512
learning_rate=0.02,max_depth=10, metric='binary_logloss',min_data=10, num_leaves=40,objective='binary', sub_feature=0.5	0.779009823947087

Table-4.4 Hard churn LightGBM classifier results

Conclusion

Model Optimisation for Customer Churn

To sum up, our project effectively tackled the difficult problem of anticipating customer attrition in the agriculture industry by differentiating between hard and soft attrition for UPL. Through navigating the complexities of a dataset containing more than 230 features, we produced noteworthy results. The project reduced the dimensionality and streamlined the dataset, which improved the performance of the model. By using several machine learning techniques, we were able to forecast both soft and hard churn with greater accuracy. Working together with domain specialists guaranteed the interpretability and usefulness of the model, enabling well-informed decision-making. The model's usefulness was increased by optimising for scalability and computing efficiency. Additionally, by creating a structure for ongoing refinement and enhancement, the initiative acknowledged the fluid character of the agriculture sector and consumer conduct. Ultimately, this project produced a churn prediction model that is reliable, accurate, and flexible. It is well-suited to offer insightful analysis and guide tactical customer retention initiatives in the dynamic agricultural business environment.

Automation Of Cross Sell Impact Calculation Pipeline

The transition from manual lead matching to an advanced, data-driven, automated cross-sell pipeline during this internship assignment at UPL has been nothing short of revolutionary. Through the use of sophisticated machine learning models, thorough hyperparameter tuning, and in-depth feature importance analysis, the project's goals of increasing operational efficiency, improving accuracy, and fostering a more personalised approach to cross-selling were met. In addition, the creation of a dynamic dashboard gave us the ability to see the results of the project and connected us in real time to the delta table, guaranteeing that the insights were up to date and easily accessed. The project's dedication to data-driven decision-making and efficient operations was highlighted by this integration.

References

[1] LightGBM: A Highly Efficient Gradient Boosting Decision Tree (neurips.cc)

[2] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. Annals of statistics, pages 1189–1232, 2001

[3] Charles Dubout and François Fleuret. Boosting with maximum adaptive sampling. In Advances in Neural Information Processing Systems, pages 1332–1340, 2011.

[4] Manuel Fern´andez-Delgado, Eva Cernadas, Sen´en Barro, and DinaniAmorim. Do we need hundreds of classifiers to solve real world classifi-cation problems? Journal of Machine Learning Research, 15:3133–3181,2014

[5] Amit, Y. & Geman, D. (1997). Shape quantization and recognition with randomized trees. Neural Computation, 9, 1545–1588.

[6] Freund, Y. & Schapire, R. (1996). Experiments with a new boosting algorithm, Machine Learning: Proceedings of the Thirteenth International Conference, 148–156.

[7] M.I. Jordan et al. Machine learning: trends, perspectives, and prospects Science (2015)

[8] J. T. Hancock and T. M. Khosghoftaar, "CatBoost for Big Data: An Interdisciplinary Review," Journal of Big Data, vol. 7, no. 94, pp. 1-45, 2020.