Web Application for Churn Prediction

By

Khandoker Zishan 012211013

Submitted in partial fulfilment of the requirements
of the degree of Master of Science in Computer Science and Engineering

March 22, 2022

Department of Computer Science and Engineering
United International University

Abstract

Customers are the driving force for a business to thrive. To make a new product, decisions have to be made whether this is something which the customers will resonate with. This is why it is absolutely necessary to understand a customer and whether or not they will stay with the company. Businesses capture many information of a customer which at a glance will not make much sense. But those information are key features for building a model. This is where the web application for Churn Prediction will step in which will take in all the various information about a customer and predict whether that customer will churn or not. Simply uploading an excel file with customer data will show which customers will churn or not. This will allow business managers to intervene to keep those potential customers from churning.

Acknowledgements

This work would have not been possible without the input and support of many people over the last two trimesters. We would like to express my gratitude to everyone who contributed to it in some way or other.

First, we would like to thank my academic advisor, Dr. Swakkhar Shatabda for his constant guidance and support.

We are also thankful to Rifat for providing us with real company data to work on.

Last but not the least, we owe to our family including our parents for their unconditional love and immense emotional support.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter discusses churn in a broader aspect and how they can be classified. The chapter then proceeds to talk about the motivation, objectives, methodology and the structure of the report.

1.1 Problem Statement

Churn prediction is one of the most common problem of industries which comes under the consumer behavior category in data mining. Making new customers is five times costlier than retaining of old customers. Churn is a vast field. It is not limited to an organization or in telecommunication industries. Churn is also a major concern in wireless sensor network. Stephan Holzer proposed a concept for churn prediction in wireless sensor networks. He introduced nodes which observed newly joined or recently left nodes in wireless network. Churn is classified into two categories: Voluntary and Involuntary churn (Figure 1.1). Fraud cases, non-payment cases or device/services are not in use are example of Involuntary churn. Whereas voluntary churn is classified in two categories. When customer moves from one place to another, deliberate change in technology or psychological change of circumstances incidental churn take place. When company try its level best to resolve the movement of customers, but customers don't agree.

1.2 Motivation

While there are many companies that cater to building churn prediction models, there are several barriers which restrict all companies availing services from those companies. Therefore, there is a demand for custom application, and this is exactly the aim of this project.

1.3 Objectives

The project aims to achieve the following objectives:-

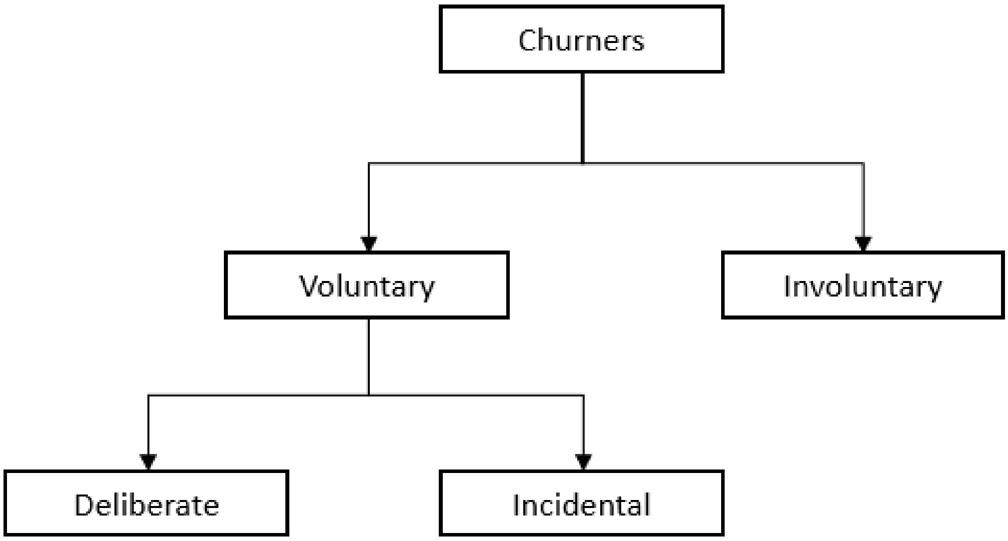


Figure 1.1: Classification of Churn

1. To develop a supervised machine learning model to predict churn from a given set of inputs.
2. To develop a web application for customer churn prediction.
3. To perform comparative analysis of algorithms and feature engineering for customer churn prediction.

1.4 Methodology

This project is done using the customer data set of 12,000 users gathered from a delivery logistics company. The data set was derived from the database server using BigQuery. Data was cleaned to remove duplicates. Descriptive analysis was performed on the data set to identify the important features which impacted customer's decision to churn. Different models were trained on the selected features.

1.5 Overview of the report

The report is composed of 5 chapters including the Introduction. Other chapters are briefly described below:

1. Chapter 2 discusses the background of the project along with a literature review.
2. Chapter 3 describes the dataset and how explanatory analysis was performed on the data. Machine learning algorithms used and how they were evaluated along with their results.

3. Chapter 4 discusses about the development of the web application with system designs, software architecture, implementation details and also a demonstration of the application in action via case analysis.

4. Chapter 5 summarizes what the project has achieved and what can be further done to improve it along with the limitations.

Chapter 2

Background

In this chapter we will discuss about the literature review and a brief descriptions of the model.

2.1 Preliminaries

For churn prediction, 3 models were selected to train the data and make prediction. The 3 models are described below.

Decision Tree Model

Although it may be used to solve both classification and regression issues, it is more often utilized for classification. In a tree-structured classifier, each leaf node represents a result, with internal nodes representing data set attributes, branching representing decision rules. The Decision Node and the Leaf Node are the two nodes of a Decision Tree. Leaf nodes, on the other hand, are the result of such choices and do not have any more branches to follow them. Decisions or tests are made depending on the characteristics of the provided dataset [1]. The CART algorithm is used in the decision trees (Classification and Regression Trees). In both circumstances, judgments are made based on any of the features' conditions. The conditions are represented by the internal nodes, and the choice is made based on those criteria via the leaf nodes. The cost function of a decision tree can be described by

$$J(k, t_k) = \frac{m_{left}}{m} I_{left} + \frac{m_{right}}{m} I_{right}$$

where $_{left}$ is the impurity for the left and right path in the tree, $_{node}$ is the number of samples on either side on either the left or the right side of the sub-tree, and m is the total number of samples in the sub-tree [2].

Support Vector Machines

A supervised machine learning model known as a support vector machine (SVM) employs classification methods to solve classification issues involving two groups. It is possible to train an SVM model to classify fresh text after providing it with training data for each category. They offer two key benefits over newer algorithms like neural networks: faster speed and greater performance with a smaller sample size (in the thousands) [3]. A hyperplane is a (n - 1)-dimensional subspace of an n-dimensional space. The hyperplane of a two-dimensional space is a one-dimensional line. a 2-dimensional plane that slices through a 3-dimensional cube will be called a hyperplane.

$$beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + ... + \beta_n * x_n = 0$$

Any hyperplane can be written mathematically as above.
Equation for seperate hyperplane can be written as:

$$y * (\beta_0 + \beta_1 * x_1 + \beta_2 * x_2) > 0$$

The margin is the distance from one side of the dashed line to the other of the solid line. Soft Margin and Kernel Tricks are two techniques introduced by SVM to deal with non-linearly separable scenarios. By repurposing already-existing features and transforming them, Kernel Trick generates brand-new functionality. In order for SVM to locate the nonlinear decision border, it will need these additional characteristics.

Artificial Neural Network

It's a hardware and/or software system modeled after the way neurons work in the human brain. Artificial neural networks (ANNs), often known as deep learning or artificial intelligence (AI), are a kind of deep learning technology that also comes under this umbrella. Complex signal processing and pattern recognition challenges tend to be the focus of commercial applications of these technologies. Handwriting recognition for check processing, speech-to-text transcription, oil-exploration data analysis, weather prediction, and face identification are only a few examples of commercially important uses since 2000 [4]. Examples of significant commercial applications since 2000 include handwriting recognition for check processing, speech-to-text transcription, oil-exploration data analysis, weather prediction and facial recognition. [5].This form of machine learning algorithm is based on the human brain, and is known as an Artificial Neural Network (ANN). Learning from example data sets is a significant benefit of ANN. The most typical use of an ANN is to approximate a random function. It is possible to arrive at solutions that specify the distribution in a cost-effective manner using these kinds of technologies. The inputs/information from the outside world are delivered to the model in the Input Layer, which is also known as the Input Nodes. The information is passed on to the Hidden layer from the Input nodes. This layer of neurons is where all calculations are done on the incoming data, and it is called the "hidden layer." A neural network may have any number of hidden layers. The simplest network has just one layer of hidden complexity. The model's results/conclusions are found in the model's output layer. The output layer might have one or more nodes. For binary classification, the output node is one, while for multi-class classification the output nodes might be several. The hidden layer receives data in the form of input units, each of which has a weight connected to it. Neurons are found in every buried layer. Each neuron receives information from all of its neighbors. All of the calculation is done in the hidden layer after the inputs have been passed on. All the inputs are multiplied by their respective weights in the first step. All variables are weighted according to their gradient or coefficient. It demonstrates the potency of a certain piece of information. A bias variable is inserted after allocating the weights. In order for the model to match the data as well as possible, it must have some bias. Prior to being sent to the next layer of neurons, the input is transformed using a nonlinear activation function. Every hidden layer is passed through by the algorithm before it goes to the final layer and gives us our final result. Back It is via the process of propagation that the model's error, i.e. the discrepancy between actual and anticipated values, is kept to a minimum. Optimizers are used to keep the weights current. There are a variety of ways to optimize neural networks, such as changing the weights, to reduce the amount of error. One of the optimizers used to arrive at new weights is Gradient Descent.

$$Z_1 = W_1 * In_1 + W_2 * In_2 + ... + W_n * In_n + b$$

W are the weights assigned to the inputs In and b is the bias.
the activation function is applied to the linear equation Z_1 The activation function is a nonlinear transformation that is applied to the input before sending it to the next layer of neurons. After passing through every hidden layer, the algorithm moves to the last layer to give us a final output. Back Propagation is the process of updating and finding the optimal values of weights or coefficients which helps the model to minimize the error i.e difference between the actual and predicted values. The weights are updated with the help of optimizers. Optimizers are the methods/ mathematical formulations to change the attributes of neural networks i.e weights to minimizer the error. Gradient Descent is one of the optimizers which helps in calculating the new weights.

2.2 Literature Review

If companies want to obtain maximum profits, they must divert their focus to customers instead of making new products [6] . To increase the loyalty of customers, the process of Customer Relationship Management (CRM) is used [7]. A small change in customer's preference may take away all the profits of a company which is why CRM's primary focus is to address customer pain points in the most effective way possible [8]. When customer becomes interested in a different product's features, they stop using the product of a company and this scenerio is defined as customer churn. This forces the profit of a company to plummet while at the same time, the company spend more to acquire new customers [9]. Almost one-fourth of their customers are lost on average per year by many companies [10] and sometimes they even go up to 36 per cent [11]. On the contrary, revenues will climb by almost 25 to 80 per cent if per cent reduction in customer churn is lowered by 5 per cent [12]. From the above percentages, the gravity of the role customer relationship management can be understood [13]. Thus industries neglecting this will surely lose their hold in the market [14].

2.3 Summary

In essence, this chapter tries to describe the 3 algorithms used to train the churn prediction model. Insights from other papers have been highlighted about customer churn and how they are a vital problem that any organization have to deal with.

Chapter 3

Methodology

In this chapter, we provide where was the training data sourced from, the description of the dataset, data analysis and the parameters for the machine learning models and how they are evaluated and finally the outcome.

3.1 Description of Dataset

The dataset contained over 12000 rows of data and had 74 columns gathered from a delivery logistics company. The data was derived from a database server using BigQuery. This is structured data in a csv file format. Duplicate values were removed from the dataset and all null values were replaced with 0. The dataset is highly unbalanced as shown in Figure 3.1. The models were trained on this unbalanced dataset.

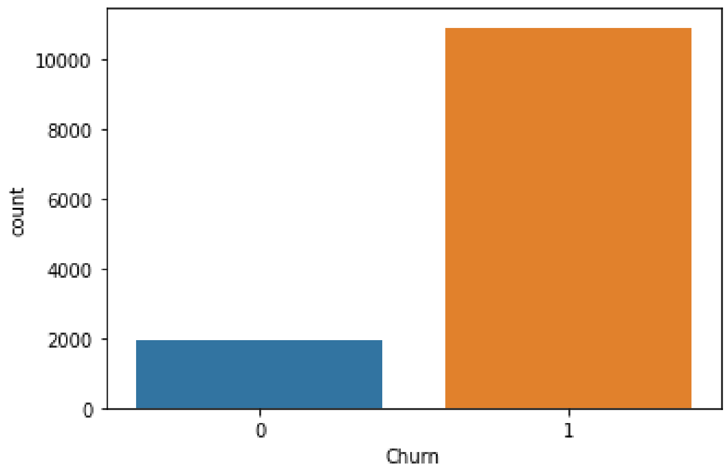


Figure 3.1: Highly unbalanced data-set

Description of each feature name is explained below:

- **Churn** - This is the label used to identify if the customer has churned. 1 denotes the customer has churned and 0 means the customer has not churned.

Feature Name	Data Type
Churn	Integer
ISD_Success	Integer
SUB_Success	Integer
OSD_Success	Integer
ConversionISD	Float
ConversionSUB	Float
ConversionOSD	Float
E2E_TAT_OSD	Float
Return_connect_ratio	Float
Return_connect_ratio_isd	Float
Return_connect_ratio_sub	Float
Return_connect_ratio_osd	Float
Distinct_merchant_booked_parcel	Integer
Success_Rate	Float
Average_attempt	Float

Table 3.1: Features with their data type

- **ISD,SUB,OSD** - These are the number of parcels delivered to each of the zones. ISD stands for INSIDE DHAKA, SUB stands for SUBURB Areas, OSD stands for Outside Dhaka.
- **Success** - Successfully delivered the parcel to the customer.
- **Conversion** - Ratio of successful deliveries made to the customer with respect to failed deliveries.
- **E2E TAT** - End-to-End TAT is the measurement in days of how long it took for the parcel to be delivered to its destination from the moment it was picked up.
- **Return Connect Ratio** - Ratio of parcel items returned back to the company.
- **Distinct merchant booked parcel** - Merchant ID who booked to deliver the parcel.
- **Success Rate** - Success rate of the merchant.
- **Average Attempt** - Number of times the parcel took to finally get delivered to the customer.

3.2 Explanatory Data Analysis

Since there are 74 columns, feature analysis was performed to carve out the features which impacted churn the most. A heatmap was generated and from there only the highest correlated features were considered and this reduced the columns to 16.

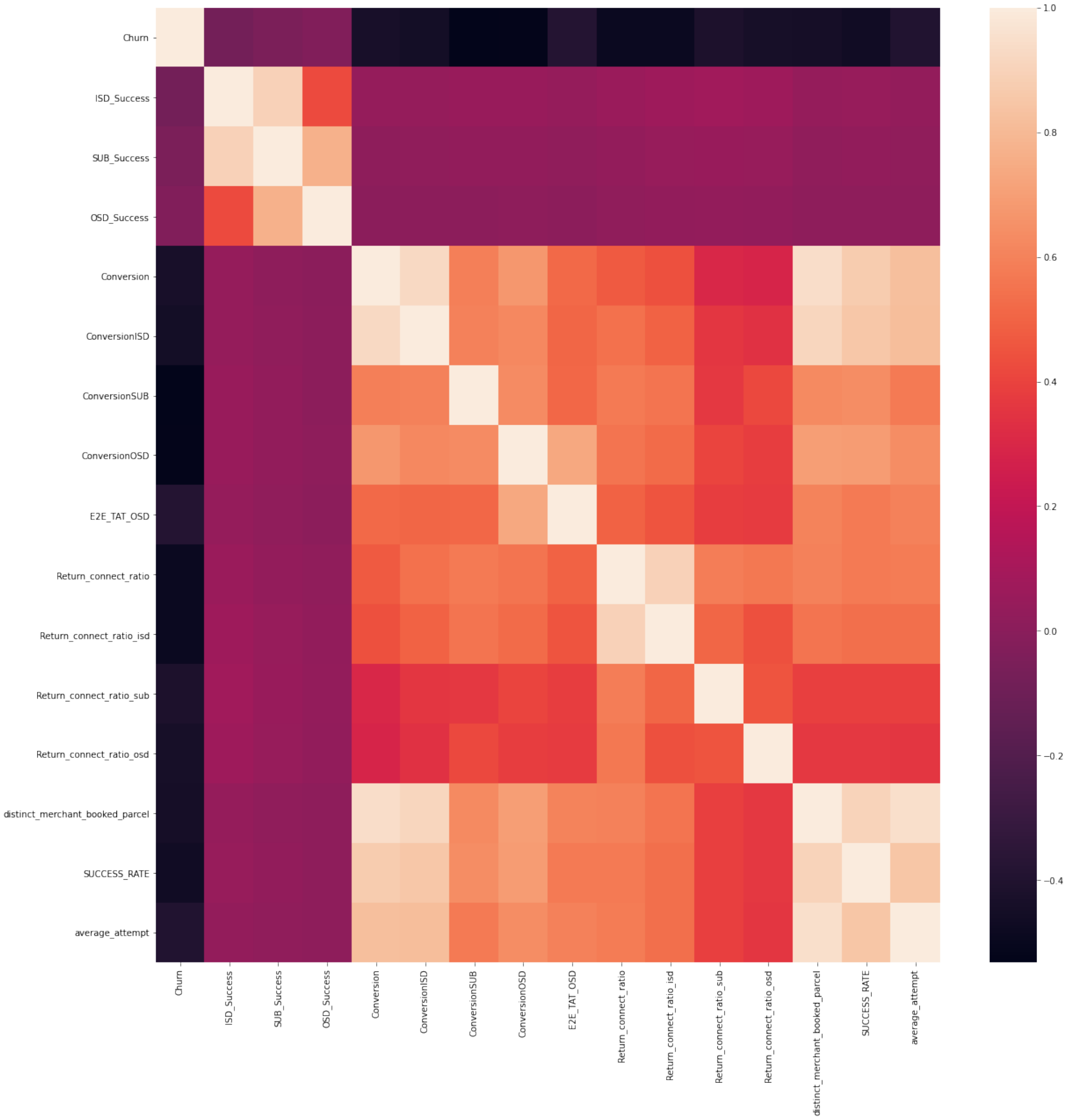


Figure 3.2: Heatmap to show correlation of features with respect to Churn

3.3 Machine Learning Algorithms

To keep the comparison between the models fair, all the models were trained based on 70-30, where 70 per-cent of the data went into training the model and the rest on testing the model. Decision tree classifier model used from the Scikitlearn library. Support vector classifier model has also been used from the Sciklitlearn library. The artificial neural network was trained with a total of 5 layers and since it is a classification problem, sigmoid was used at the final layer for predicting if the customer will churn or not. 100 epochs was selected for the training.

3.4 Evaluation

The metrics used for evaluation are described below.

3.4.1 Confusion Matrix

In the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a matching matrix. It is a table with 4 different combinations of predicted and actual values. True Positive are values that are predicted to be positive and they are positive. True negative are values that are predicted negative and they are true. False Positive (also known as Type 1 Error) is when the value predicted is positive and it is false. False Negative (also known as Type 2 Error) is when the value is predicted negative and it is false.

		Actual Values	
		Positive	**Negative**
Predicted Values	**Positive**	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative(TN)

Table 3.2: Confusion Matrix

3.4.2 Accuracy

It is the fraction of predictions the model got right.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

3.4.3 Precision

Precision is the fraction of true positive examples among the examples that the model classified as positive. In other words, the number of true positives divided by the number of false positives plus true positives.

$$Precision = \frac{TP}{TP + FP}$$

3.4.4 Recall

Recall, also known as sensitivity, is the fraction of examples classified as positive, among the total number of positive examples. In other words, the number of true positives divided by the number of true positives plus false negatives.

$$Recall = \frac{TP}{TP + FN}$$

3.4.5 F1-score

The F-score, also called the F1-score, is a measure of a model's accuracy on a dataset. It is used to evaluate binary classification systems, which classify examples into 'positive' or 'negative'.

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

3.4.6 Support

Support is the number of actual occurrences of the class in the specified dataset.

3.5 Results

After training all the models, the ANN and SVM models had an accuracy of 90 per-cent and the DTM was not far behind at 88 per-cent and the results are captured in Table 3.3. The confusion matrix results for all the models are shown. SVM has the least Type

Prediction Model	**Prediction Accuracy**
DTM Churn Prediction Model	88%
SVM Churn Prediction Model	90%
ANN Churn Prediction Model	90%

Table 3.3: Accuracy comparison between the models

2 Error amongst the 3 models at 72. ANN model has the least Tpe 1 Error at 210.

	Predicted churn customers	**Predicted unchurn customers**
Actual churn customers	356	210
Actual unchurn customers	164	3129

Table 3.4: Confusion Matrix for ANN Model

	Predicted churn customers	**Predicted unchurn customers**
Actual churn customers	356	259
Actual unchurn customers	215	3027

Table 3.5: Confusion Matrix for DTM

	Predicted churn customers	**Predicted unchurn customers**
Actual churn customers	244	331
Actual unchurn customers	72	3212

Table 3.6: Confusion Matrix for SVM

		Precision	Recall	F1-score	Support
ANN	**0**	0.68	0.63	0.66	566
	1	0.94	0.95	0.94	3293
DTM	**0**	0.62	0.58	0.60	617
	1	0.92	0.93	0.93	3242
SVM	**0**	0.77	0.42	0.55	575
	1	0.91	0.98	0.94	3284

Table 3.7: Precision, Recall, F1-score and Support evaluation for the models

ANN model has the highest precision amongst the 3 models at 0.94 for predicting Churn, however SVM model has the better precision for non-churn values amongst the 3. SVM has the highest true positives for churn values as shown by the Recall column. However, ANN has the highest true positives for non-churn values. F1 score for churn values are equal for ANN and SVM however for non-churn values, ANN edges out the SVM. From the support column, we can see the data was highly imbalanced.

From all the tables, it can summarized that ANN model does the best even though ANN and SVM share the same prediction accuracy, but when it comes to F1 score, ANN beats out SVM model to predict whether a customer will churn or not. ANN is the best model to use.

Chapter 4

Application Development

In this chapter, we discuss about the software architecture and the requirements of software needed to complete the web application. We also show step by step method on how the application works.

4.1 System Design

Data Flow Diagram (DFD) Level 0 and 1 are shown below.

4.1.1 DFD Level 0

The user will input data into the Web App in the form of a csv file. The file will be processed by the app, and it will display the prediction back to the user.

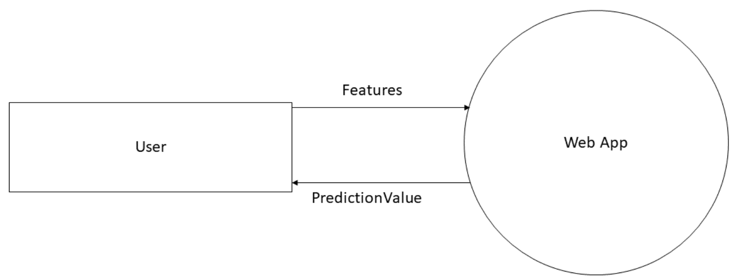


Figure 4.1: Data Flow Diagram Level 0

4.1.2 DFD Level 1

Once the data is input into the web app, it will first convert the data, in Dataframe converter process, into a dataframe which can be processed by the predictive models. Once the conversion has happened, based on the user's selection of the model, the data will be sent to the .h5 file of any of the 3 models. In the model, the data will be processed to get a value for prediction. The predicted value will be sent to the web app which will display the result in a human readable format for the user.

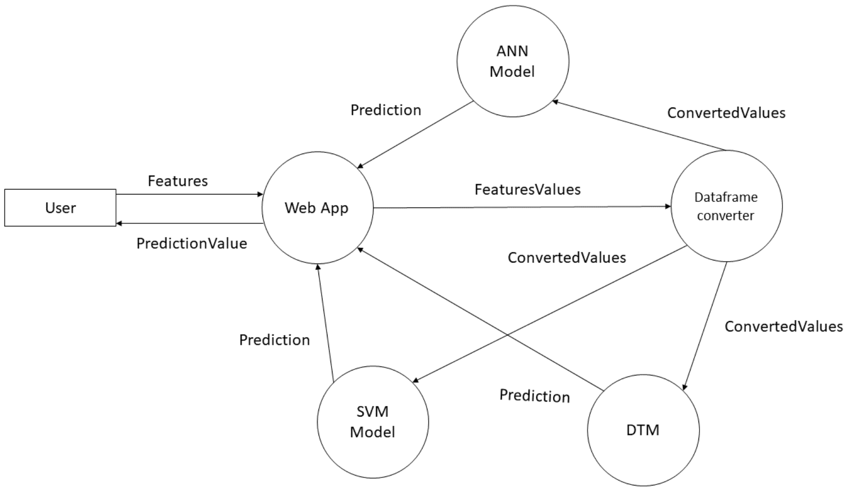


Figure 4.2: Data Flow Diagram Level 1

4.2 Software Architecture

The following section details the software architecture required for this app's development as shown in Figure 4.3. Frontend: Flask, a python-based web framework, is used to create the UI. Flask is chosen due to its simplicity and flexibility. It will capture the input of the user as well display the prediction. Backend: Python is used as well as the models are created based on the packages of python. The backend will store all the 3 models as h5 files and will be fed the proper data and will give the prediction.

4.3 Implementation Details

The following software were required to complete the project:

- Anaconda Navigator - version 2.0.4 [ifhardwaredoesnotincludeaNvidiaGPU]
- PyCharm Community - version 2021.1.3
- Python 3.8
- TensorFlow-2.3.0

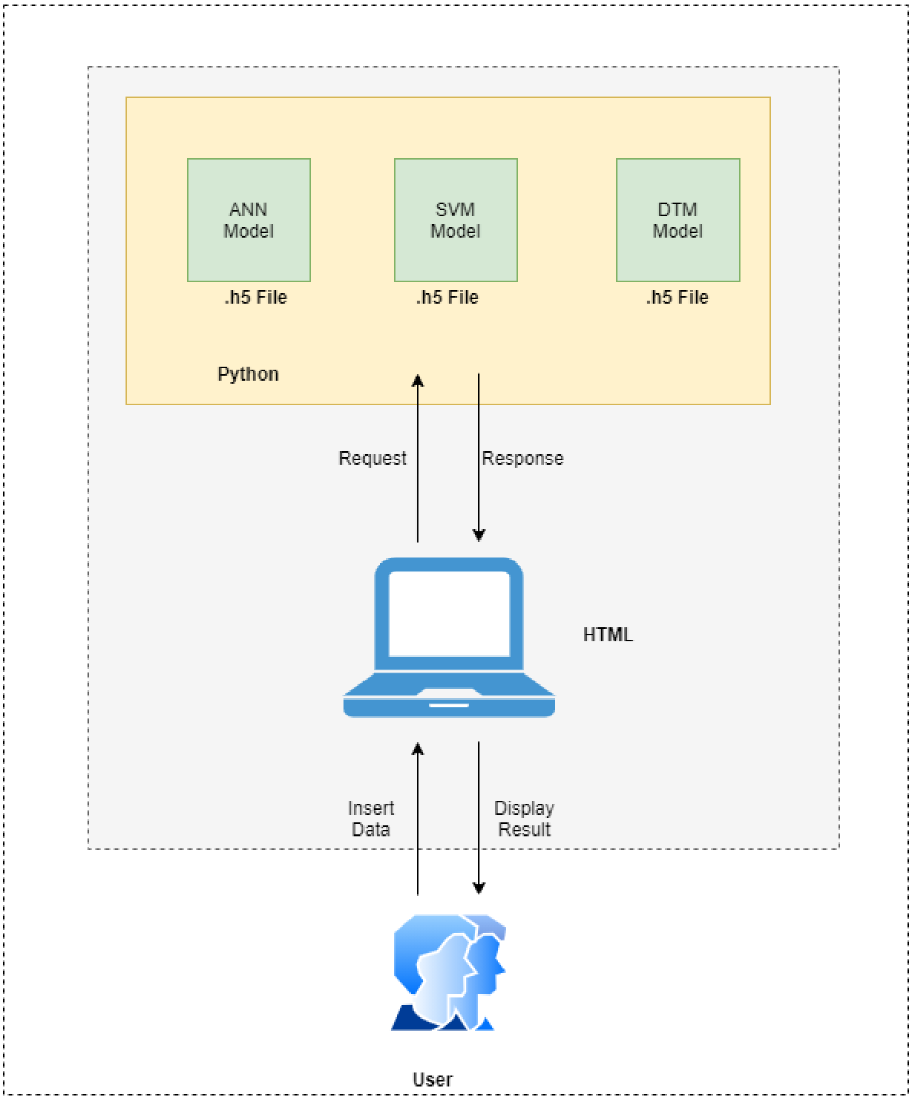


Figure 4.3: Software Architecture

- Flask-1.1.2
- Pandas - 1.3.0
- Scikitleam - 0.24.2
- Google Chrome- version 92.0.4515.107

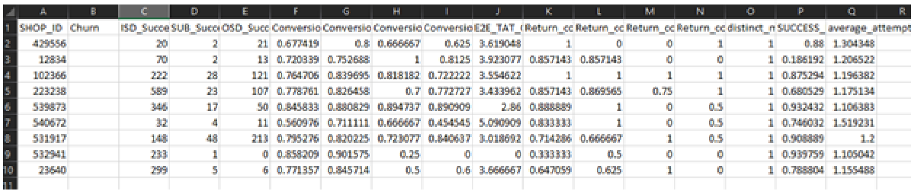
4.4 Case Analysis

For this case, the following dataset (Figure 4.4) will be imported by the web application.

After launching the application, the following screen will show up as Figure 4.5.

Import the customer data file and select the model, in this case, ANN is selected as shown in Figure 4.6.

Click on predict and the following table will be generated as shown in Figure 4.7.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	SHOP_ID	Churn	ISD_Succe	SUB_Succ	OSD_Succ	Conversio	Conversio	Conversio	Conversio	E2E_TAT_(	Return_cc	Return_cc	Return_cc	Return_cc	distinct_n	SUCCESS_	average_attempt	
2	429556		20	2	21	0.677419	0.8	0.666667	0.625	3.619048	1	0	0	1	1	0.88	1.304348	
3	12834		70	2	13	0.720339	0.752688	1	0.8125	3.923077	0.857143	0.857143	0	0	1	0.186192	1.206522	
4	102366		222	28	121	0.764706	0.839695	0.818182	0.722222	3.554622	1	1	1	1	1	0.875294	1.196382	
5	223238		589	23	107	0.778761	0.826458	0.7	0.772727	3.433962	0.857143	0.869565	0.75	1	1	0.680529	1.175134	
6	539873		346	17	50	0.845833	0.880829	0.894737	0.890909	2.86	0.888889	1	0	0.5	1	0.932432	1.106383	
7	540672		32	4	11	0.560976	0.711111	0.666667	0.454545	5.090909	0.833333	1	0	0.5	1	0.746032	1.519231	
8	531917		148	48	213	0.795276	0.820225	0.723077	0.840637	3.018692	0.714286	0.666667	1	0.5	1	0.908889	1.2	
9	532941		233	1	0	0.858209	0.901575	0.25	0	0	0.333333	0.5	0	0	1	0.939759	1.105042	
10	23640		299	5	6	0.771357	0.845714	0.5	0.6	3.666667	0.647059	0.625	1	0	1	0.788804	1.155488	
11																		

Figure 4.4: Test data-set image

Figure 4.5: Webpage of the application

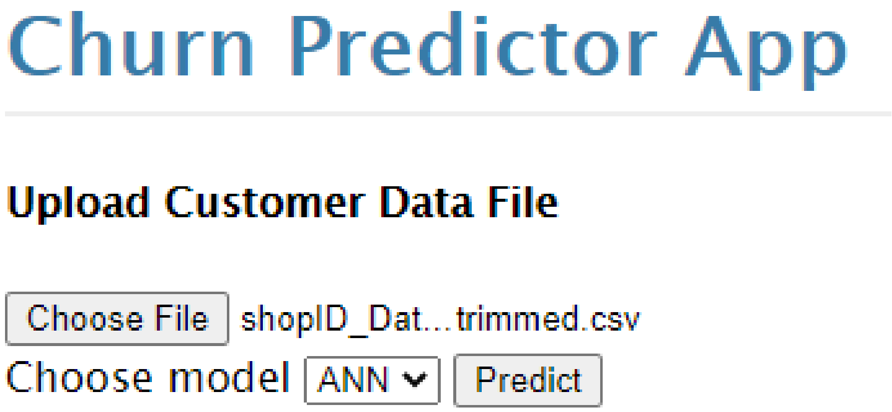
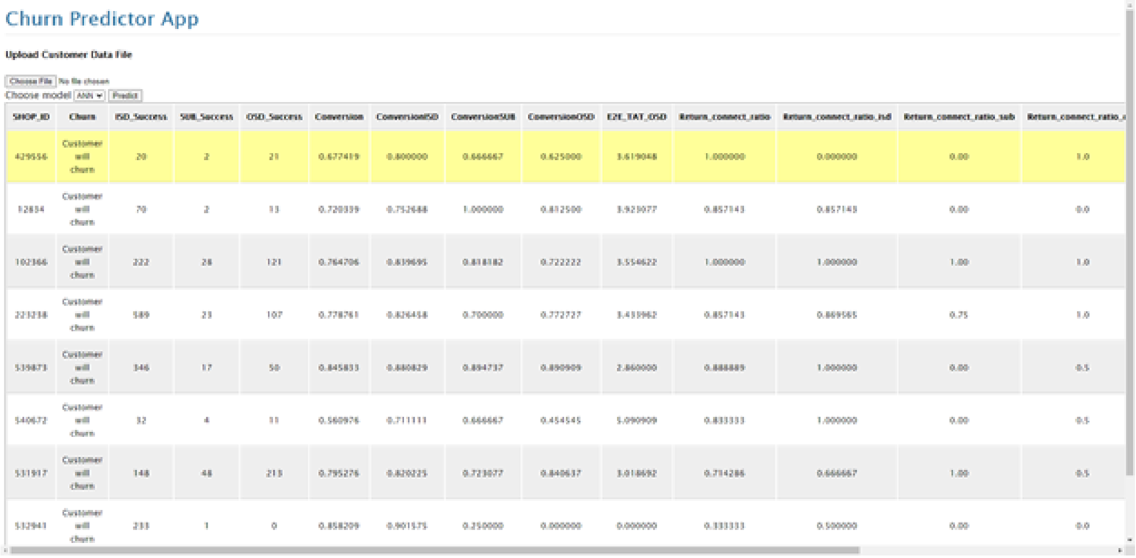


Figure 4.6: CSV file uploaded at the application

4.5 Summary

This application is simple to use as the manager only must upload the data from excel directly to the web application and select which model and predict the churn. The application also allows for predicting multiple customer data simultaneously so all the customer prediction can be seen at glance.



SHOP_ID	Churn	ISD_Success	SUB_Success	OSD_Success	Conversion	ConversionISD	ConversionSUB	ConversionOSD	E2E_TAT_OSD	Return_connect_ratio	Return_connect_ratio_isd	Return_connect_ratio_sub	Return_connect_ratio_
429556	Customer will churn	20	2	21	0.677419	0.800000	0.666667	0.625000	3.619048	1.000000	0.000000	0.00	1.0
12834	Customer will churn	70	2	13	0.720339	0.752688	1.000000	0.812500	3.923077	0.857143	0.857143	0.00	0.0
102366	Customer will churn	222	28	121	0.764706	0.839695	0.818182	0.722222	3.554622	1.000000	1.000000	1.00	1.0
223238	Customer will churn	589	23	107	0.778761	0.826458	0.700000	0.772727	3.433962	0.857143	0.869565	0.75	1.0
539873	Customer will churn	346	17	50	0.845833	0.880829	0.894737	0.890909	2.860000	0.888889	1.000000	0.00	0.5
540672	Customer will churn	32	4	11	0.560976	0.711111	0.666667	0.454545	5.090909	0.833333	1.000000	0.00	0.5
531917	Customer will churn	148	48	213	0.795276	0.820225	0.723077	0.840637	3.018692	0.714286	0.666667	1.00	0.5
532941	Customer will churn	233	1	0	0.858209	0.901575	0.250000	0.000000	0.000000	0.333333	0.500000	0.00	0.0

Figure 4.7: Table generated with prediction of customer churn

Chapter 5

Conclusion

This chapter summarizes about the web application and the limitation of it and what future work can be done to make the web application even better.

5.1 Summary

From the evaluation, ANN model is the recommended one to be used. This project allows managers to obtain important information about the customers and learn why a customer churn. By understanding those features, the managers can take decisions to reduce churn percentage. Plus knowing that a customer is on the verge of churning, the manager can attend to that customer and offer services so that the customer does not churn.

5.2 Limitation

The model is trained on a specific set of features and thus, if there are new data points captured which has strong correlation to the churn, then the whole model has to be retrained and also codes need to be adjusted so that the web application can accommodate for it.

5.3 Future Work

For future work, the data can be stored in a database, so that the predicted values can be compared against the real outcome of the customer and these data can be used to further train the model.

References

[1] Mrinal Pandey and Vivek Kumar Sharma. A decision tree algorithm pertaining to the student performance analysis and prediction. *International Journal of Computer Applications*, 61(13), 2013.

[2] Xiangfei Meng, Pei Zhang, Yan Xu, and Hua Xie. Construction of decision tree based on c4. 5 algorithm for online voltage stability assessment. *International Journal of Electrical Power & Energy Systems*, 118:105793, 2020.

[3] Shan Suthaharan. Support vector machine. In *Machine learning models and algorithms for big data classification*, pages 207–235. Springer, 2016.

[4] KP Sudheer, AK Gosain, D Mohana Rangan, and SM Saheb. Modelling evaporation using an artificial neural network algorithm. *Hydrological Processes*, 16(16):3189–3202, 2002.

[5] Jon Kleinberg and Eva Tardos. *Algorithm design*. Pearson Education India, 2006.

[6] Rosalind McMullan and Audrey Gilmore. The conceptual development of customer loyalty measurement: A proposed scale. *Journal of Targeting, Measurement and Analysis for Marketing*, 11(3):230–243, 2003.

[7] Judith W Kincaid. *Customer relationship management: getting it right!* Prentice Hall Professional, 2003.

[8] Rajshree Agarwal, Raj Echambadi, April M Franco, and Mitrabarun B Sarkar. Knowledge transfer through inheritance: Spin-out generation, development, and survival. *Academy of Management journal*, 47(4):501–522, 2004.

[9] Kristof Coussement and Dirk Van den Poel. Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. *Expert Systems with Applications*, 36(3):6127–6134, 2009.

[10] Wei-yu Kevin Chiang, Dilip Chhajed, and James D Hess. Direct marketing, indirect profits: A strategic analysis of dual-channel supply-chain design. *Management science*, 49(1):1–20, 2003.

[11] Aslıhan Dursun and Meltem Caber. Using data mining techniques for profiling profitable hotel customers: An application of rfm analysis. *Tourism management perspectives*, 18:153–160, 2016.

[12] Frederick F Reichheld and W Earl Sasser. Zero defeofions: Quoliiy comes to services. *Harvard business review*, 68(5):105–111, 1990.

[13] D Van der Poel and B Larivière. Investigating the role of product features in preventing customer churn, by using survival analysis and choice modelling: the case of financial services. *Expert Systems with Applications*, 27(2):277–265, 2004.

[14] Guo en Xia and Wei dong Jin. Customer churn prediction model based on svm. *Systems Engineering-Theory and Practice*, pages 71–77, 2008.