

A BI Tool for the AI based BI Framework for Churn Prediction of the Telco Industry and Getting the Prescriptive Decisions by Minimizing the Churning Customers

Rohan Samarasinghe¹ and Hansa Perera²

^{1,2} Department of Information Technology,
Faculty of Technology,
University of Colombo,

¹rohan.ict.cmb.ac.lk, ²hansa.deva.perera@gmail.com

Abstract. Information has become an integral in both individuals and the business and a large amount of digital information in new types is being generated every moment from different data capturing devices. As today we are living in an on-demand on-command world, the data analytics is playing an increasingly critical role for making full use of it in most of the business organizations. Different forms of analytics can be found, but they don't provide the best perspective or provide the right information. To turn a company into a profitable operation, top management relies heavily on decisions, which are in turn based on the type of knowledge available. Hence, we are going to introduce a new AI based BI Framework which can be used by different businesses, industries and service providing companies for predictions and to get the prescriptive decisions. In this paper, we are going to introduce a Business Intelligence tool for the AI based BI framework to analyse the big data set and to extract the required data and create a predictive model for making the prescriptive decisions for the business. This BI tool is sophisticated, user-friendly, cost-effective and comprehensive data-driven business intelligence tool for efficient prediction, and it can be used to analyse the big data sets of the industry. The BI tool must be trained using a data set of the same industry before it is used with the large data set. The tool first extracts the data set, transform the extracted data set to the required format and then it is loaded to the BI tool to select the suitable algorithm. The tool has a set of pre-defined algorithms and the loaded data set is analysed using these algorithms and get the percentage of accuracy in each algorithm we introduced with the tool. Then the end-users will be able to select the suitable algorithm for the data set and then it can be used with the big data set of the similar industry. The AI based BI framework provides the facility for many cross industries and businesses. In this work, we focused on the Telco Industry and we implemented our novel Business Intelligence tool for that industry. Because they are facing problems in finding hidden patterns in their large chunk of data which may have a big impact on their decision-making process. Most telecommunication companies are suffering from churn (customer leaving the service) of their customers day by day due to high competition within the competitor telco companies. This churning will badly effect to the huge financial losses and survival of such companies. The experiment result showed more than 90% accuracy of selecting the right churn customers and the tool helped to make the prescriptive decisions to the top management to minimize churning.

Keywords: Data-mining, Descriptive analytic, Predictive analytic, Perspective analytic, Statistical Analysis, Visual Analytics, Business Intelligence

1 Introduction

Information has become more integral to the business organizations to take critical decision which will directly influence to the survival of the business[1]. With the proliferation of new devices, introducing new data capturing techniques, development of ICT technology and cheaper the storage devices directly correlate to increase the massive or to big data set of the business organizations. Hence, most of the business organizations nowadays generate massive or big datasets which may have the main characteristics of variety, velocity, veracity and volume. When the size of the data is getting bigger and unstructured, the complexity of the data set will increase and generally accepted data analytical techniques[2],[3] cannot be used with the big data sets. Hence, special techniques and tools must be used with such big data set.

To get such critical decisions, there are four types of data analytics that would improve the decision-making process. As shown in figure 1, these analytic types are Descriptive Analytics, Diagnostic Analytics, Predictive Analytics and Prescriptive Analytics[4] .

Descriptive analytics answers the question of what happened and juggles raw data from multiple data sources to give valuable insights into the past. In diagnostic Analytics, historical data can be measured against other data to answer the question of why something happened. These two types of techniques can be used easily with the small structured primary data sets getting high accuracy.

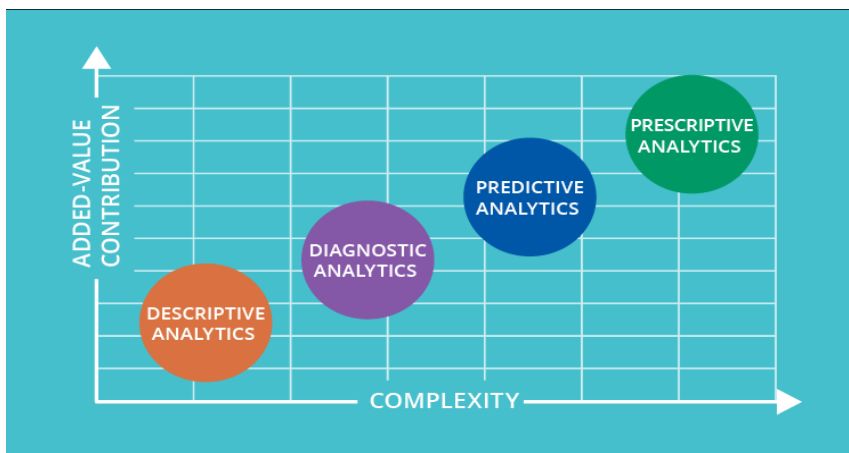


Fig. 1. Four types of Data Analytic Techniques

Predictive analytics which tells that what happened to the business and why that happened, what is happening now and what is likely to happen in future.

What is likely to happen is a branch of data mining, which combines business knowledge and statistical analytical techniques, to extract hidden predictive information and to business question that needs to be answered. The purpose of prescriptive analytics is to literally prescribe *what action to be taken* to eliminate the future problem or take full advantage of a promising trend.

Our Research, we are going to introduce AI based BI Framework that can be used in any industry and business organizations to make a predictive analytic of the big dataset of the business organization and the framework proposes prescriptive analytics to literally prescribe *what action to be taken* to eliminate the future problem or take full advantage of a promising trend. Figure 2 shows the framework with the components.

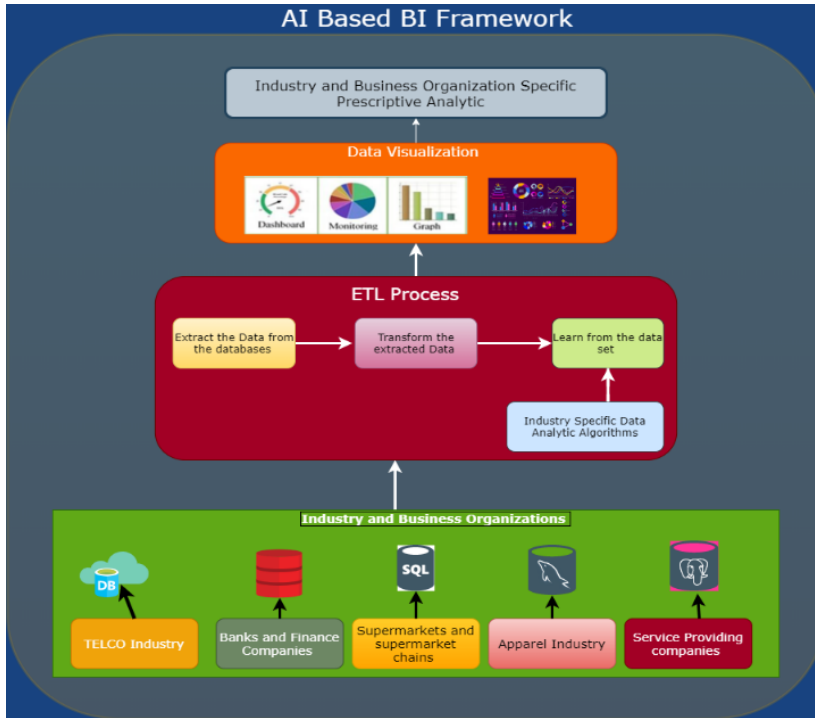


Fig. 2. AI based BI Framework

1.1 Components of the AI based BI Framework.

Cross Industry and Business Organizations

The framework can be used by any industry which produce a massive or bigdata sets for predictive analytic. Here, the industry and business organizations can be categorised as Telecommunication Companies (Telco Companies), Production and Manufacturing, Banks and financial companies, Apparel industry and Service Providing Companies. Each of these industry and business organizations has a set of **Use Cases**, then the organizations can easily decide what they want to analyse. For instance, Default Churn Prediction is one Use Case of the telco companies and Default loan Prediction also a Use Case of the Banks and Financial Companies.

ETL (Extract, Transform and Load) Process

When the business organization must select the right industry type and then select the Use Case that they want to make predictive analytics. The data may have different formats and sometime whole big dataset is unstructured. Data must be extracted and transform to the required format and load to the system. Here, in this frame work we use different Machine Learning algorithms [5], [6] [2] to analyse the data. Figure 3 shows algorithms that can be used with the use cases we introduce in this framework.

When the algorithms are selected for the use cases we mainly consider the size, quality, and nature of data; the available computational time; the urgency of the task; and the purpose of the business organizations.

Data Visualization

Data Visualization is the first step to make sense of data and to get the decisions by the top management. The data must be translated and presented and provide the right data correlations in a simple way using a wide range of techniques — charts, diagrams, maps, etc. Choosing the right technique and its setup is often the only way to make data understandable. Vice versa, poorly selected tactics won't let to unlock the full potential of data or even make it irrelevant.



Fig. 3. Machine Learning Algorithms

Here, in this work we use Python language for the ETL process and using the libraries of Python the data visualization will be done.

Industry and Business Organization Specific Perspective Analysis

Based on the predictive analytic, the framework proposes what actions to be taken to eliminate the future problem. For instance; if the telco company wants to minimize the churning customers from the system, they can propose the free data packages or free SMS package to those customers.

The data must be analysed periodically to see the perspective analytics are working properly for the industry.

We regularly train the components of the AI based BI framework with use case and update the industry specific algorithms with new algorithms and industry specific perspective analytic components with new suggestions.

I this work we introduce a Use Cases for the Telco Companies and Banking sectors to analyse the call records (big dataset) of the customers and identify the churn customers and analyse the data of the customers who get the loans from the banks. This Churn Prediction BI tool is sophisticated, user-friendly, cost-effective and comprehensive data-driven business intelligence tool that can be used to analyse the big data set by those who do not have deep knowledge in data science. The Churn Prediction BI tool aims to create a collection of resources, technologies, and programmed products that can be used to gather,

incorporate, aggregate, and make data accessible for better, quicker, and more reliable decisions in the telecommunications and financial industries for early prediction of churn customers since consumer churn is a big issue in telecommunications companies.

According to research [7], telecommunication firms have the highest churn rate of all companies, ranging from 10-67 percent. Customers may quit for a variety of reasons, but customer frustration is the most influential factor. Customers exhibit such churning behaviors that could be used as a parameter to predict whether or not they would leave the service.

Customers being loan defaulters has been a big problem in the finance and banking sectors. According to the research conducted for this article [8], a customer's inability to repay a bank loan results in a significant financial loss for the banking and finance industry. As a result, we created a research product that offers valuable, completely data-driven, and fact-based information and analysis to top-level management in companies in advance, allowing them to make effective, accurate, and result-oriented strategic decisions. ViviSight allows the forecast more accurate by using statistical analysis and data mining to measure the likelihood of each customer leaving the service or defaulting on their loan.

2 Methodology

We used the agile approach to carry out this research project. Before beginning the study, we conducted an investigation to determine the major business challenges that the companies were facing. After defining problem domains, we conducted a literature review of available business intelligence tools, noticing their shortcomings and shortfalls. We then created ViSight, a solution that solves the major business issues.

We also conducted a conventional feasibility analysis, demonstrating that ViviSight is financially, technically, and operationally feasible due to its open-source architecture and lack of technological constraints and dependencies. Then we concentrated on gathering ViviSight's functional and non-functional criteria.

During the design process, we created a high-level architecture design to integrate all of ViviSight's functional and non-functional specifications.

We also moduled the system's basic workflow and architectural design. ViviSight is made up of two parts: a desktop application and a web-based application. Architecture is classified into three groups. 1.ETL 2. Predictive model building and Evaluation 3. The dashboard. The high-level architecture of ViviSight is depicted in Figure 4.

2.1 Architecture by ViviSight

2.1.1 ETL (Extract, Transform and Load):

Extract: In a real-world environment, different data sources related to the device can be stored in various locations and in various formats.

For Billing Support System, Operation Support System, Customer Relationship Management, Customer Experience Management, and so on, there may be several sources. These data can also be in a number of formats, such as xls, xlsx, csv, txt, xml, and so on. As required, the system should be able to classify these data sources.

When considering a data warehouse, we can use Oracle, SQL, MySQL, or another technology. The system should be capable of extracting data from any format. The DWH should be used to pick and extract the required attributes for the prediction model.

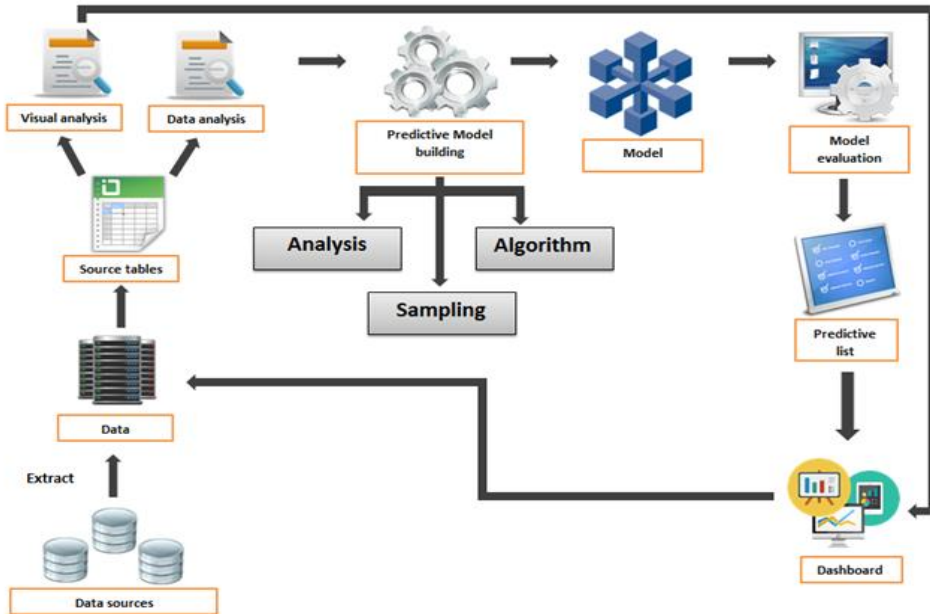


Fig. 4. System Architecture of ViviSight

Transform

We used a collection of rules to convert the data from the source to the target in the transform stage. We performed a variety of operations, including joining data from multiple sources, generating aggregates, generating surrogate keys, sorting, calculating new values, and applying advanced validation rules.

Load

The loading phase of the ETL process is the final step. Tables are used to load and store data from various data sources.

Loading is carried out for two reasons:

1. **Visual Analytics:** This component was used to display the transformed data in a graphical format (bar charts, pie charts, tables, etc.) so that the user could analyze it and get a sense of how current and historical data behaved. Various types of reports are created based on the user's requirements. Users can track any changes in the data pattern if they are visible.
2. **Data Analytics:** In this component we store the data in a form of tables in order to use in predictive analytics. Predictive models use these tables as input for the analytics process.

2.1.2 Predictive Model Building and Evaluation

In this process, we have used most-effective approaches in predictive analytics to create a statistical model of future behaviour of given data set. In predictive analytics, we have used classification approach of data mining concerned with forecasting probabilities and trends. We have developed various algorithms using classification techniques because we are dealing with binary classification problems: customer will be churn or not and customer will be loan

default or not. As a result, we used data mining classification techniques in our model building process.

We used confusion matrix techniques to verify the accuracy of constructed algorithms during the Model Evaluation process.

2.1.3 Dashboard

Dashboard is a web-based component in ViviSight that visualizes data in a content-rich manner and provides a high-level perception of data insights and facts. We've created graphical reports that explain data in the most effective way possible. ViviSight Dashboard provides advanced data visualization and guided analysis through auto charting, and offers an array of visualization techniques to present data and results in the most insightful way.

Implementation

Before directly going into the implementation, we have designed the basic workflow of our system that helped us in an effective implementation of ViviSight.

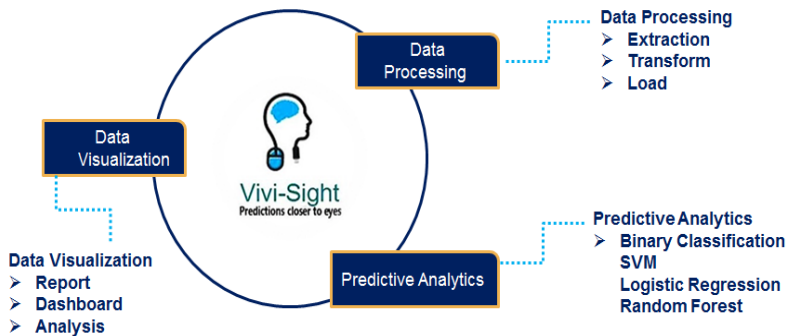


Fig. 5. Basic Workflow of ViviSight

We organized our activities for the effective implementation of ViviSight into the following stages, taking into account both functional and non-functional requirements.

- Data preprocessing
- Identifying the most significant data attribute
- Data is sampled into two sets: training and test set
- Building a model
- When possible, benchmarking each model and comparing it to others to determine which is the most accurate.
- Creating a Dashboard
- Integration with the ViviSight System

The first and most significant move in our research was to acquire the dataset. We were able to obtain a data set from N-able, a Sri Lankan business intelligence firm. The Telco data set contains 10,000 data instances with various attributes, while the Finance data set contains 20,000 data instances.

To deal with missing data values, we used data imputation techniques. We discovered that for those data instances with errors during data entry by users, we need to use data

imputation techniques based on the provided dataset. We also discovered that some data fields were left blank and that some values contained irrelevant data. We used the median value to fill in for missing values. To adapt it to the models, we also removed the data field that has the same values across all data instances.

Table 1. Description of important data attribute for finance-dataset

Parameter	Description
loan amount	Amount of loan requested by borrower
int_rate	Interest rate of the loan
annual_inc	Annual income of borrower
purpose	Main intension of borrower for asking a loan
installment	If a loan is issued, the borrower must make a monthly payment.
sub-grade	More details about the loan
total_acc	The total number of credit lines in the borrower's credit line at the moment.
emp_length	Employment details of borrower
desc	Length of loan description

After preprocessing the financial data collection, Table 1 displays a list of useful data attributes with their descriptions.

Table 2. Description of important data attribute for Telco-dataset

Parameter	Description
network_stay	Network usage duration.
outgoing_call_duration	Time usage for the outgoing calls.
weekly_revenue	Weekly Income.
no_of_complains	Total number of complaints last 3 months.
no_of_short_calls	Number of missed calls.
average_recharge_amount	Usual recharge amount.
average_incoming_minutes	Usual incoming call duration.
last_month_no_of_sms	Number of short messages last month.
active_value_added_services	Services receive other than calls.

After preprocessing the Telco data set, Table 2 shows the list of useful data attributes and their descriptions.

We were now concerned about the data sampling for the training and test sets. We divided the data into two categories: training data (75%) for feeding models and testing data (25%) for benchmarking and determining the most accurate model.

For statistical modeling and programming, we next chose between Python and R.

Table 3. Comparison of R and Python

R	Python
High Complexity	Comparatively low complex than R
Already written packages for statistical modelling	Best if major concern is programming
Syntactically less cleaner than Python	Syntactically more cleaner than R
The documentation for the libraries isn't very user-friendly.	Rich documentation for libraries
It's more difficult to incorporate into a manufacturing workflow.	It's also simple to incorporate into a production workflow.
It comes with a large set of libraries for various statistical analyses.	For machine-learning activities, the scikit-learn library, panda, scipy, and numpy are fantastic.

Python was chosen as the most appropriate programming language for ViviSight for the following reasons:

- ETL and Dashboard are included in ViviSight, and they will be incorporated with the statistical model building section. As a result, Python is used because it is simple to incorporate into a production workflow.
- Python is easier to code than R from a programming standpoint.
- Python includes libraries for machine learning, such as scikit-learn.
- Python's pandas package has matured to the point that it can now outperform R dataframes, especially in terms of time-series indexing.

We concentrated on developing predictive models using data-mining and machine learning techniques after deciding on a programming language.

We divided our prediction model building into two parts since we are dealing with a classification problem: whether the customer will churn or not, and whether the customer will default on their loan.

1. Churn Prediction Model
2. Loan Default Prediction Model

To build these prediction models, we used the classification algorithm below.

a. Random Forest

A forest of trees is grown in this process, and variation is introduced among the trees by projecting the training data into a randomly chosen subspace before fitting each tree [3]. During the learning process, it builds the tree at random. Every leaf node has an internal node where the attributes are evaluated and a classification is assigned.

If B is the number of trees in the Random Forest

For $b = 1, \dots, B$:

Sample, with replacement, n training examples from X, Y ; call these X_b, Y_b

We trained a decision tree f_b on X_b, Y_b .

After training, predictions for unseen samples x' [3] can be made by averaging the predictions from all the individual regression trees on x' :

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B \hat{f}_b(x') \quad (1.1)$$

b. Support Vector Machine

Support vector machines have recently become popular for solving classification problems. We map the data into a higher-dimensional input space and create an optimal separating hyperplane in that space using this method. We have studied the research work on SVM by (Vapnik, 1995; Vapnik, 1998a; Vapnik, 1998b) and have used the same notation as Christopher Bishop's book "Pattern Recognition and Machine Learning" to make it easier to read. [4] The sign of $y(\mathbf{x})$ can be used to determine the class of an input vector in SVM.

$$y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (1.2)$$

If $y(\mathbf{x}) > 0$ we assign \mathbf{X} to class +1 and if $y(\mathbf{x}) < 0$, we assign it to class -1. In this case, $\phi(\mathbf{x})$ is a feature-space transformation that can map \mathbf{X} to a space with higher, potentially infinite, dimensions.

c. K-nearest Neighbors

It is used in the Classification style method to give weight to the contributions of the neighbors, so that the neighbors who are closer contribute more to the average than those who are farther away.

A KNN estimate is the average of the outcomes of the k-nearest neighbors.

$$y = \frac{1}{K} \sum_{i=1}^K y_i \tag{1.3}$$

Where y_i is i^{th} case of the examples sample and y is the prediction (outcome of query point). Odd values of $y = 1, 3, 5$ are used in binary classification tasks to prevent ties, i.e., two classes labels achieving the same score. Odd values of $y = 1, 3, 5$ are used in binary classification tasks to prevent ties, which occur when two labels in the same class receive the same score.

d. Logistic Regression

A classification model is a type of logistic regression. We try to predict the outcome of the independent variables in classification models, which can be categorical or numerical.

$$p_{\alpha}[y|x] = \frac{1}{Z(x)} \exp(y\alpha \cdot \Phi(x)) \tag{1.4}$$

We used the training set to train the prediction models after they were developed using the following approaches. We used the following approach to benchmark the prediction models we had built so far after they were trained using the training set.

i. Confusion Matrix as Performance Metrics

	Predicted Negative	Predicted Positive
Actual Negative	TN	FP
Actual Positive	FN	TP

Fig. 6. Confusion Matrix as Performance Metrics

A confusion matrix is often used to determine a classifier:

Where TN denotes the number of correctly classified negative examples, FN denotes the number of incorrectly classified positive examples, FP denotes the number of incorrectly classified negative examples, and TP denotes the number of correctly classified positive examples.

ii. Receiver Operating Curve (ROC)

The Receiver Operating Characteristic (ROC) curve is a popular performance metrics between true positive and false positive error rates. For a ROC curve, the Area Under Curve (AUC) is a commonly used and agreed output metric.

The ROC curve depicts a binary classifier's results. The X-axis represents the false positive rate ($=FP/(FP+TN)$), while the Y-axis represents the true positive rate ($=TP/(TP+FN)$). The optimal threshold point can be found through locating the point on ROC that is closest to (0, 1).

We concentrated on implementing two arms of the ViviSight application after selecting the above classification techniques in order to construct a churn prediction model and a loan default prediction model.

a. Desktop Application

The ETL and Model Building components of ViviSight are included in our desktop application. It was built with Jython libraries, which gave us the freedom to create attractive user interfaces. The key user interface for the desktop application is depicted in Figure 7.



Fig. 7. Home screen for Desktop application

b. Web - based application

Another important part of ViviSight is the Dashboard. It's used to display data in a visual way. There are two ways to visualize data. One method is to use exploratory data, while the other is to use insightful data. For graphical reports and visual analytics, it was built with javascript, php, and high-charts. The dashboard configuration of ViviSight is shown in Figure 8.



Fig. 8. Home screen for Web application

3 Research Findings, Results and Evidence

The ETL component, prediction outcome, model accuracy, and Dashboard features of the final ViviSight system were all benchmarked. This section displays the results.

ViviSight ETL component has unique feature of mapping given dataset attribute to the database fields. It will help to proper transformation of dataset attribute to acceptable fields.

We used scikit-learn, a Python machine learning library, to create our models for this study. We have used `sklearn.ensemble`. For the Random Forest model, and `sklearn.linear` model we used `RandomForestClassifier`. `LogisticRegression` for Logistic Regression, `sklearn.svm.SVC` for Support Vector Machine, `sklearn.neighbors.KNeighborsClassifier` for K-Nearest Neighbors.

In the given data set, we were able to determine the likelihood of each consumer churning. Figures 9 and 10 display probability calculations based on various algorithms.

ACCOUNT_ID	Gender	Age	Probability_rf	Probability_svm	Probability_knn
10617598	Male	65	0.5	0.4	0.6
10620056	Female	20	0.2	0.1	0
10620266	Female	74	0.1	0.2	0

Fig. 9. Churning probability calculated using different model

Figure 9 shows how we saved the churning likelihood that we measured using various algorithms into a database, which then served as the backend for our data visualization for predictive analytics in Dashboard.

Probability_rf= probability calculated by Random Forest

Probability_svm= probability calculated by SVM

Probability_knn=probability calculated by K-NearestNeighbor

id	loan_amnt	int_rate	Probability_rf	Probability_svm	Probability_lr
631535	7000	13.80%	0.3	0.1	0.3
632417	35000	20.25%	0.4	0.6	0.7
633105	16000	19.74%	0.3	0.2	0.4

Fig. 10. Loan-default probability calculated using different model

We have stored the loan default likelihood that we estimated using various algorithms in table 4.

Probability_lr= probability calculated by logistic regression

We chose the most powerful and commonly used models for forecasting churn in the telecommunications industry and loan default in the banking and finance industry as part of our study. We used ROC-AUC Values as the output matrix to benchmark the model so far produced.

Table 4. ROC-AUC values for telco-churn prediction models

Prediction Model	AUC Values
Random Forest	0.855
Support Vector Machine(SVM)	0.836
K-Nearest Neighbor	0.831

Table 4 shows that the Random Forest model is the most reliable model for churn estimation, with an AUC score of 0.855.

Table 5. ROC-AUC values for loan Default prediction models

Prediction Model	AUC Values
Logistic Regression	0.821
Support Vector Machine (SVM)	0.813
Random Forest	0.810

Table 5 shows that the Logistic Regression model is the most reliable model for churn prediction, with an AUC score of 0.821.

4 Conclusions and Future Work

This paper discusses the various methods that can be used to construct a prediction model using statistical models, data mining, and machine learning techniques. ViviSight distinguishes itself from competing products by offering a comprehensive solution that includes ETL. All-in-one prediction and dashboard.

ViviSight has been designed primarily to solve problems in the telecommunications and financial industries. On the one hand, we will concentrate on improving the accuracy of these models, maintaining the customer profiling mechanism, and improving the efficiency of the prediction process by implementing more efficient data cleaning and transformation techniques in our next step. It is hoped that the findings of this research will be useful to anyone planning to develop a similar system or any other real-time system, as they will provide insight into the efficiency, accuracy, and reliability levels that can be anticipated with the tools, technologies, and programming approach considered in this paper.

Acknowledgment

We would like to thank the Sri Lanka Institute of Information Technology (SLIIT) for providing us with a conducive atmosphere and facilities in which to complete this project. We also wanted to express our gratitude to our families and friends for their patience and encouragement in helping us finish this project.

- [1] Arthur Middleton Hughes, "Churn reduction in the telecom industry," , 24 January 2007. [Online]. Available: <http://www.dmnews.com/dataanalytics/churn-reduction-in-the-telecom-industry/article/94238>. [Accessed 20 June 2015].
- [2] Andy Liaw and Matthew Wiener, "Classification and Regression by randomForest," , 3 December 2002. [Online].

- [3] Available:
ftp://131.252.97.79/Transfer/Treg/WFRE_Articles/Liaw_02_Classification %20and%20Oregression%20by%20randomForest.pdf. [Accessed 23 March 2015].
- [4] Mathieu, "Support Vector Machines in Python," 19 September 2010. [Online]. Available: <http://www.mblondel.org/journal/2010/09/19/support-vector-machines-in-python/>. [Accessed 1 June 2015].
- [5] Jiayu Wu, "Loan Default Prediction Using Lending Club Data," 14 May 2015
- [6] Jitendra Nath Pandey, Maheshwaran Srinivasan., "Predicting Probability of Loan Default," 15 December 2011. [Online]. Available: <http://cs229.stanford.edu/proj2011/PandeySrinivasan-PredictingProbabilityOfLoanDefault.pdf>. [Accessed 2 May 2015]
- [7] Yhat, "Machine Learning for Predicting Bad Loans," 16 August 2013. [Online]. Available: <http://blog.yhathq.com/posts/machine-learning-for-predicting-bad-loans.htm>. [Accessed 22 July 2015].

References

- [1] R. H. Sampieri, *No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析 Title*.
- [2] B. Richmond, "Introduction to Data Analysis Handbook.," *Acad. Educ. Dev.*, pp. 1–27, 2006, [Online]. Available: <http://eric.ed.gov/?id=ED536788>.
- [3] S. Citation, *Strengthening Data Science Methods for Department of Defense Personnel and Readiness Missions*. 2017.
- [4] "4-types-of-data-analytics @ www.scnsoft.com." [Online]. Available: <https://www.scnsoft.com/blog/4-types-of-data-analytics>.
- [5] T. Loya and G. Carden, *Business intelligence and analytics*. 2018.
- [6] S. Guide, I. Deep, L. Models, and T. B. Ii, *Introduction to Deep Learning Using R Introduction to Deep*.
- [7] Arthur Middleton Hughes, "Churn reduction in the telecom industry," , 24 January 2007. [Online]. Available: <http://www.dmnews.com/dataanalytics/churn-reduction-in-the-telecom-industry/article/94238>. [Accessed 20 June 2015].
- [8] Roger M. Stein, "The relationship between default prediction and lending profits: Integrating ROC analysis and loan pricing," , May 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378426604000895>. [Accessed 22 June 2015].