An intelligent system for churn prediction and customer retention: a case of telecommunications company

Parth Sarangi

Computer Science and Information Management
Asian Institute of Technology Thailand

Thesis Defense, April 2018

Table of Contents

Overview

- Telecom industry is highly competitive
- Government deregulation policies
- Affordable handsets and Technological advancements
- Disruptive plans and services by rival companies
- Database reveals trends of service usage
- Data mining to identify churn customers

Overview(contd.)

Operators	Customer Count in Aug 2016	Increase or Decrease in Period				Customer Count in Dec 2016
		Aug - Sep	Sep – Oct	Oct – Nov	Nov – Dec	
Airtel	257	2	2	1	2	265
Vodafone	200	0.5	1	0.8	1.8	204
Tata Indicom	58	- 1	- 1	- 1	- 1.6	52
Reliance Jio	0	15	19	16	20	72

- TRAI - Telecom regulatory authority of India
- Reports subscribers at end of every month
- Table shows Relance Jio acquiring 72 million at end of 4 months
- Launch of 4G services by Jio have jolted the revenues of Airtel, Vodafone, Tata Indicom.
- High customer churn noticed

Problem Statement

- Constant product marketing by competitors
- Various cost effective data schemes. Night time free or high speeds, or unlimited usage plans
- Proactive mindset of incumbent services provider to identify unfaithful customers
- High investment cost of acquiring new customer
- Not a fully integrated system developed for churn prediction

Objectives

Overall objective - develop an intelligent system for churn prediction and customer retention ICPCR.
Specific objectives :

- Design models and evaluate prediction performance for churn.
- Build the system of intelligent churn prediction and customer retention system.
- Evaluate the system for reliable performance.

Limitations and Scope

- Many models for churn prediction.
- Scope of this thesis is tentatively limited to build ICPCR with 3 models - Decision tree , Support Vector Machine , Artificial Neural Network

Table of Contents

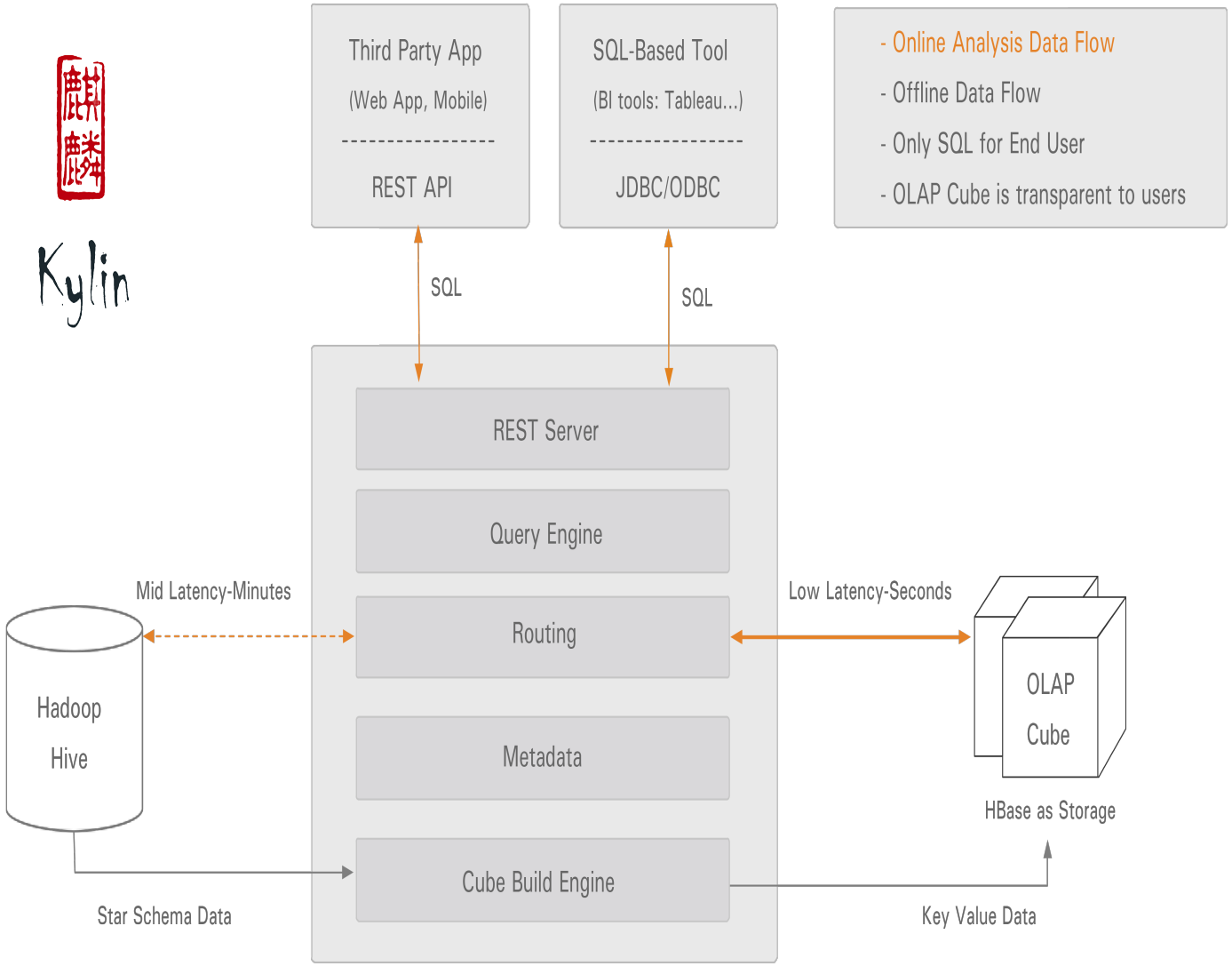
Customer Churn & Retention

- In the research paper it was found that Companies profit if they can retain customer
- Customers are most valuable asset.
- Long serving customers influence new customers to buy contracts
- Average Revenue Per User ARPU for telecom company is high if a customer stays
- If customers churn there is a loss of revenue
- Also acquiring new customer is expensive
- Forbes predicted a 10% swing in revenue if customers are retained

OLAP & Datawarehouse

- Data warehouse is a collection of Data marts
- Data marts are generally summarized tables of important data from Business units
- OLAP - Online analytical processing
- OLAP cube is the heart of an OLAP system
- There are two types - MOLAP & ROLAP. MOLAP is most common
- Apache Kylin is an open source OLAP solution

OLAP & Datawarehouse(contd.)



Data Mining

- John Naisbett (author of famous 'Megatrends') said "We are drowning in information but starved for knowledge"
- Data mining techniques can broadly be classified into two categories
  - Supervised learning
  - Un-Supervised learning

Data Mining(contd.)

Supervised Learning : The dependent and control variables are known. Classification and regression algorithms

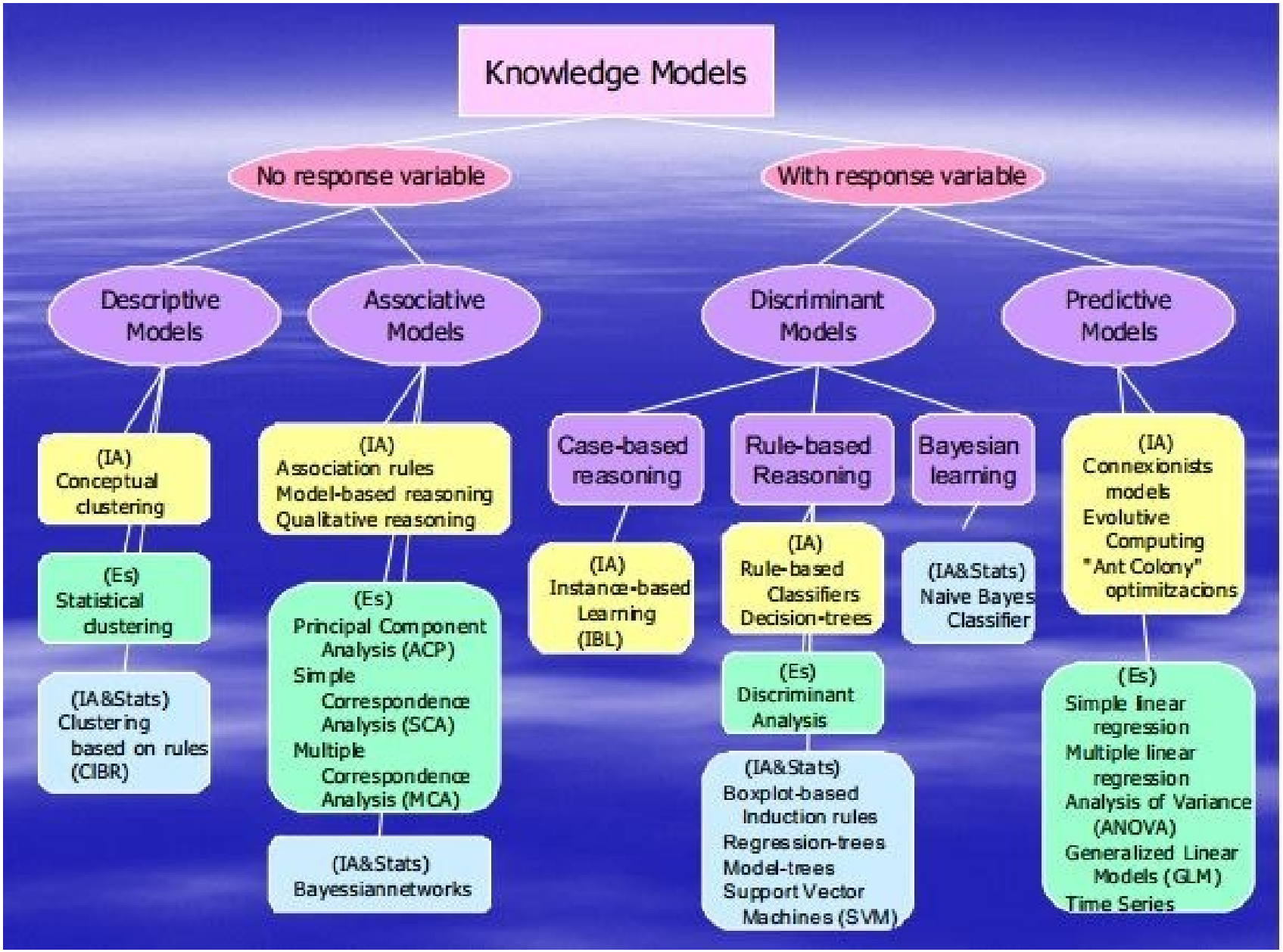
- Linear
- Multiple
- Nonlinear
- Logistic
- Decision tree
- Random forest

Data Mining(contd.)

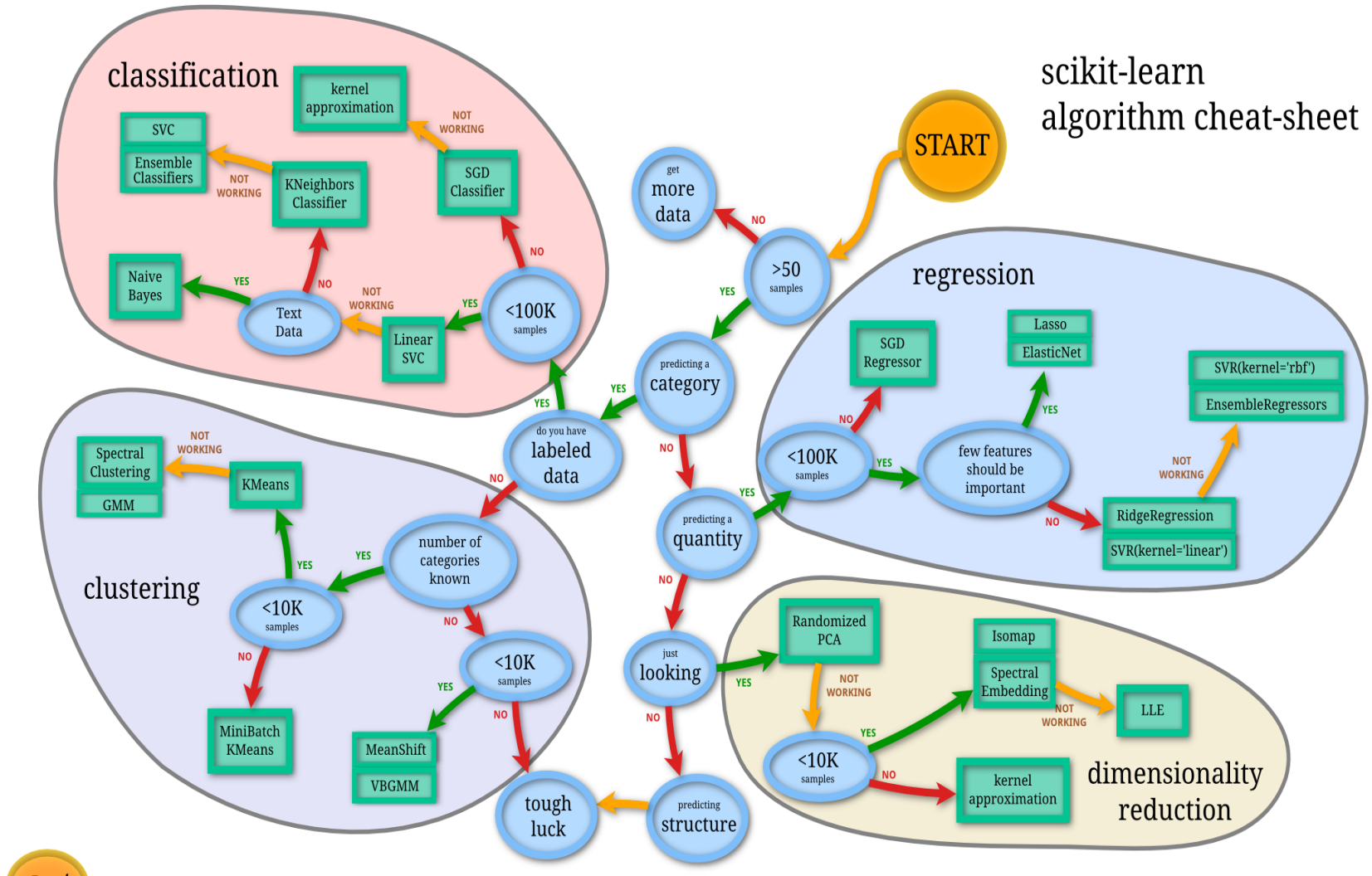
Un-Supervised Learning : The dependent and independent variables are unknown

- k means clustering
- Apriori clustering
- Hierarchical clustering
- Hidden Markov models
- Self Organizing Maps

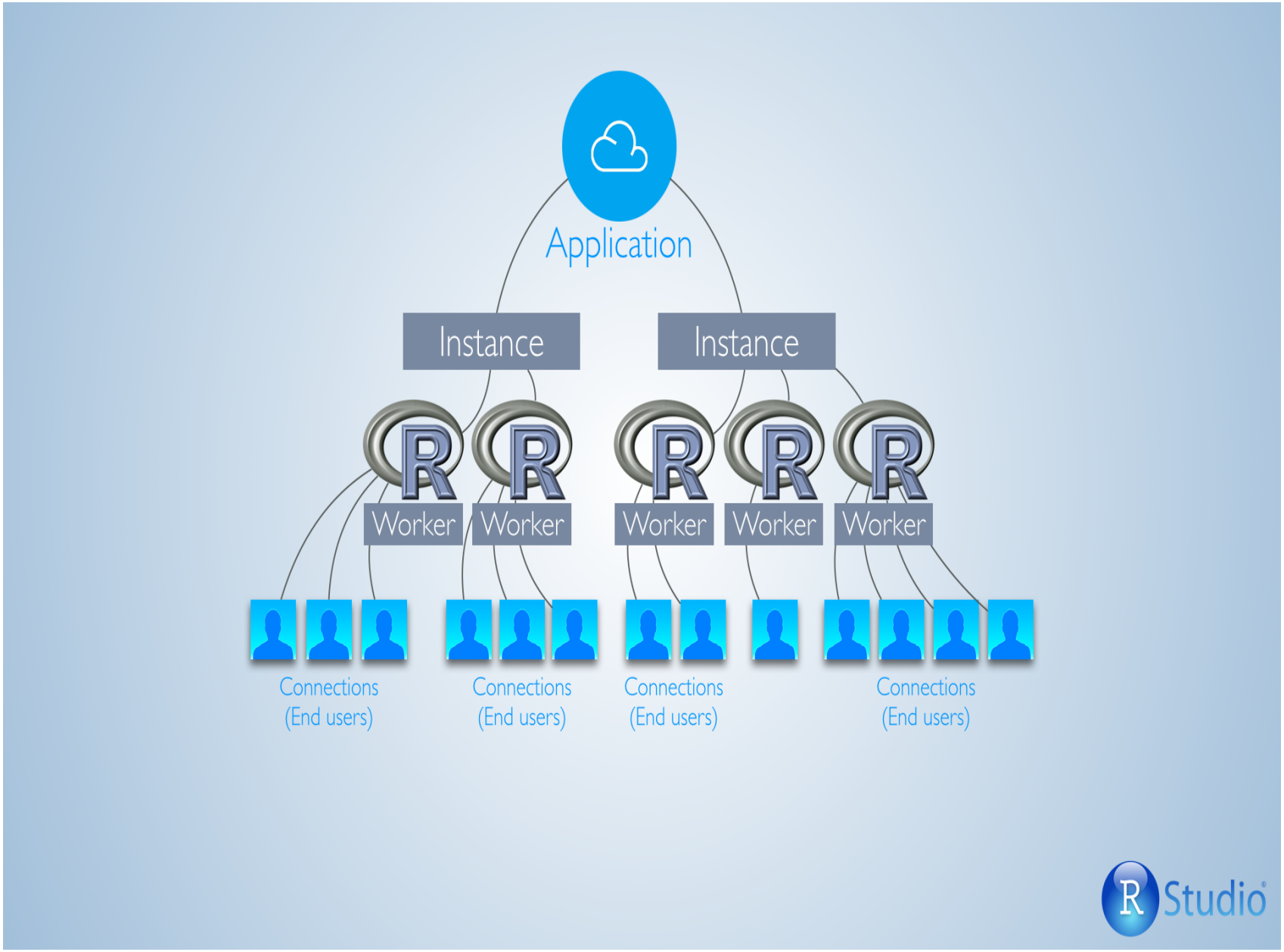
Choosing the Right Data Mining Technique



Scikit model selection



Shiny Architecture



Model Evaluation Metrics

- Holdout technique
- Cross validation technique
- Sentivity and Specificity

Review of papers - 1/5

SNo	Title & Author	Objective	Data & Methodology	Outcome	Further Research
1	Modeling & Simulation of a Predictive Customer Churn Model for Telecommunication Industry (O et al., 2015)	Adaptive neuro fuzzy inference system for prediction emulation of customer churn Neural network + fuzzy logic.	**Data :** 5000 subscribers CDR – call detail record with 21 vriables. Partitioned into 5 sets each containing 1000 records. **Method :** Number of predictor variables taken is 9. Target variable is Chrun with value Y or N. Membership function for each variable.	Found that 3 variables are very important. Total no of minute calls, no of customer service calls, no of repaired calls. Fuzzy churn model Precison 80.86% recall 92.7% and predicted accuracy 95.8%.	None suggested
2	A Hybrid Churn Prediction Model in Mobile Telecommunication Industry (Olle & Cai, 2014)	A model combined with VotedPerceptron and Logisti Regression is performance compared to the models of VP and LR as individual predictors.	**Data :** 2000 customers CDR from an Asian telecom company with 23 attributes. **Method :** A hybrid model of VP and LR was used. WEKA tool was used to model.	The hybrid model performs better than the models prediction accuracy seperately.	None suggested

Review of papers - 2/5

SNo	Title & Author	Objective	Data & Methodology	Outcome	Further Research
3	A comparison of machine learning techniques for customer churn prediction (Vafeiadis et al., 2015)	The normal model functions were performance compared to their corresponding boosted models.	**Data :** publicly hosted churn dataset at UCI machine learning repository. **Method :** Machine learning techniques of Back-Propagation algorithm , Support Vector Machines, Decision Trees, Naive Bayes and Logistic Regression were used. The boosting algorithm Adaboost.M1 a type of Adaboost was used. R progamming was used for modeling the system.	2 prediction models performed the best : 2-layer BPN with 15 hidden nodes and Decision tree classifier. SVM scored lower followed by Naive Bayes and Logit Regression at last. After application of the Boosting algo, SVM reported the best accuracy of 97% and Fmeasure over 84%.	None suggested
4	Turning telecommunications call details to churn prediction: a data mining approach (Wei & Chiu, 2002)	The company experiences a high monthly churn rate of 1.5 – 2Neural network requires a long time due to it's iterative nature. Highly skewed class distribution between churners and non-churners.	**Data :** Telecom company of Taiwan. Contractual and call details of subscribers Oct 2000 – Jan 2001. 9100000 records. **Method :** Multi classifier class combiner, Decision tree C4.5	Churn prediction is relatively high within 1 month duration. Multi classifier performs better than single classifier.	To include more variables from logs and complaints. Evaluation of empirical stats between customers from different geographic locations. Integration with datawarehouse for constantly learning behavior of customer. Research with other industry data from credit card to Internet service providers.

Review of papers - 3/5

SNo	Title & Author	Objective	Data & Methodology	Outcome	Further Research
5	Applying Fuzzy Data Mining to Telecom Churn Management (Liao & Chueh, 2011).	To determine the most effective marketing strategies of customer retention, by analyzing the responses of customers.	**Data :** Taiwan telecom company, retention activity & response data for customer contract expiry between June and Junly 2008 **Method :** ID3 decision tree for classification.	Using fuzzy set the customer retention shows that marketing via telemarketing is more effective compared with Direct mailing. Also fuzzy marketing technique is better than direct mailing marketing for customers with higher bill amounts.	Fuzzy data mining techniques to analyze the past records of results of various marketing activities to establish a marketing mode.
6	Customer churn prediction using improved balanced random forests (Xie et al., 2009).	a novel learning method, called improved balanced random forests (IBRF), and demonstrate its application to churn prediction	**Data :** Chinese bank data. 1524 [762 train, 762 test]. **Method :** IBRF = Balanced random forest + weighted random forest. Introduce 2 interval variables 'm – middle pt' & 'd – length of interval'. apply IBRF to a set of churn data in a bank as test the performance of our proposed method, we run several comparative experiments comparison of results from IBRF and other standard methods, namely artificial neural network (ANN), decision tree (DT), and CWC-SVM (Scholkopf, Platt, Shawe, Smola, & Williamson,	Accuracy rate follows this pattern $IBRF > CWC - SVM > ANN > DT$, Top-decile Lift varies as this $IBRF > CWV - SVM > DT > ANN$. IBRF offers great potential compared to traditional approaches due to its scalability, and faster training and running speeds.	Experimenting with some other weak learners in random forests. Improving effectiveness and generalization ability.

Review of papers - 4/5

SNo	Title & Author	Objective	Data & Methodology	Outcome	Further Research
7	Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques (Coussement & Poel, 2008)	Churn prediction using SVM. Benchmarked to Logit regression and random forest.	**Data :** Belgian newspaper publishing company. Training set 45000, Test set 45000 **Method :** Use of random forest software and SVM-toolbox. SVM compared to Logit regression & random forest. Grid search using 5-fold cross-validation	SVM trained on balanced distribution, outperforms logit regression when parameter selection applied. Random forest surpass SVM. Academincs and practionerx don't need to rely on traditional Logit reg, SVM with parameter selection technique and random forest offer better alternative	No complete working meta-theory to choose kernel function and SVM parameters. Thus deriving a procedure to select proper kernel function and SVM parameter.
8	Customer churn prediction by Hybrid neural networks (Tsai & Lu, 2009)	Very few studies for hybrid data mining appraoch for prediction.	**Data :** CRM dataset from American telephone company, July 2001 to Jan 2002 51,306 subscribers. **Method :** 2 methods developed and compared for performance. M1 – SOM + ANN clustering + classification is used. M2 – ANN + ANN 2 classifiers are used. 5 fold cross validation, each set of the 5 are tested 5 times. Baseline is 20 ANN's	Baseline ANN models had prediction accuracy of 88% performance : $ANN + ANN > singleANN$ $3 * 3$ SOM is best among $2 * 2$, $3 * 3$, $4 * 4$ and $5 * 5$ clustering Performance of the hybrid models is : $ANN + ANN > SOM + ANN > ANN$	Need to explore dimensionality reduction or Feature selection of data preprocessing. Application of SVM or genetic algorithms. Explore other domains for churn prediction.

Review of papers - 5/5

SNo	Title & Author	Objective	Data & Methodology	Outcome	Further Research
9	Predicting customer retention and profitability by using random forest and regression forest (Larivière & Poel, 2005)	The paper discusses more than one variable of retention and profit outcome.	**Data :** 100,000 Belgian finance company. Divided into 2 random parts, one for estimation other for evaluation. **Method :** Authors used random forest for regression to predict profitability, next purchase and defection decision. Benchmarked to linear regression model.	Random forest are better than logit and linear regression.	None suggested.
10	Churn prediction using comprehensible support vector machine: An analytical CRM application	The paper discusses more than one variable of retention and profit outcome.	**Data :** 100,000 Belgian finance company. Divided into 2 random parts, one for estimation other for evaluation. **Method :** Authors used random forest for regression to predict profitability, next purchase and defection decision. Benchmarked to linear regression model.	Random forest are better than logit and linear regression.	None suggested.
11	Churn prediction for high-value players in casual social games	Paper presents churn prediction of players of social games and the business impact of retaining high valued players.	**Data :** dataset of high value users of games - Diamond dash and Monster World, for 2 days. **Method :** The researchers trained and predicted neural networks, logistic regression, decision tree and support vector machine. Radial basis function for support vector machine was used with 10-fold cross validation. For business impact of churning the researchers designed A/B test.	Single neural network with tuned learning rate is better that other algorithms. A/B test reveals that sending free coins to high value customers does not affect churn rate.	None suggested.

Table of Contents

Research Methodology

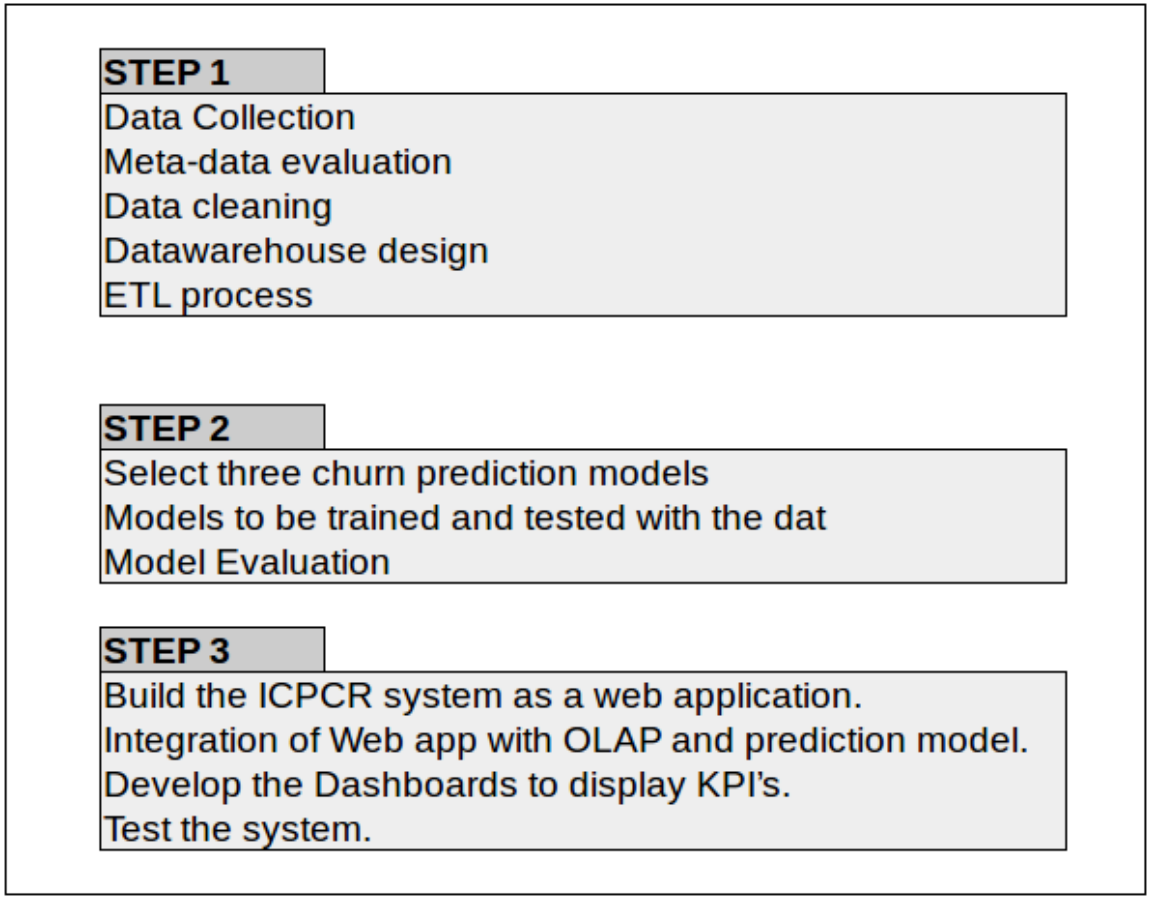


Figure: Research Methodology

Data Preprocessing and Datawarehouse Development

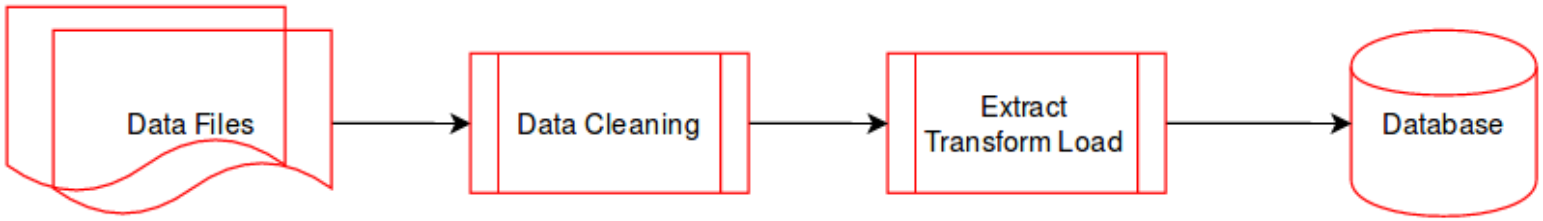


Figure: Data preprocessing

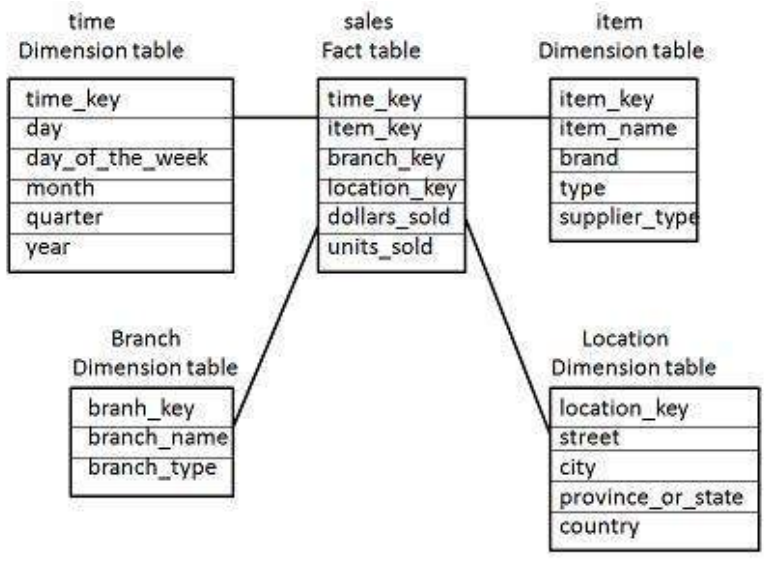


Figure: OLAP Star Schema

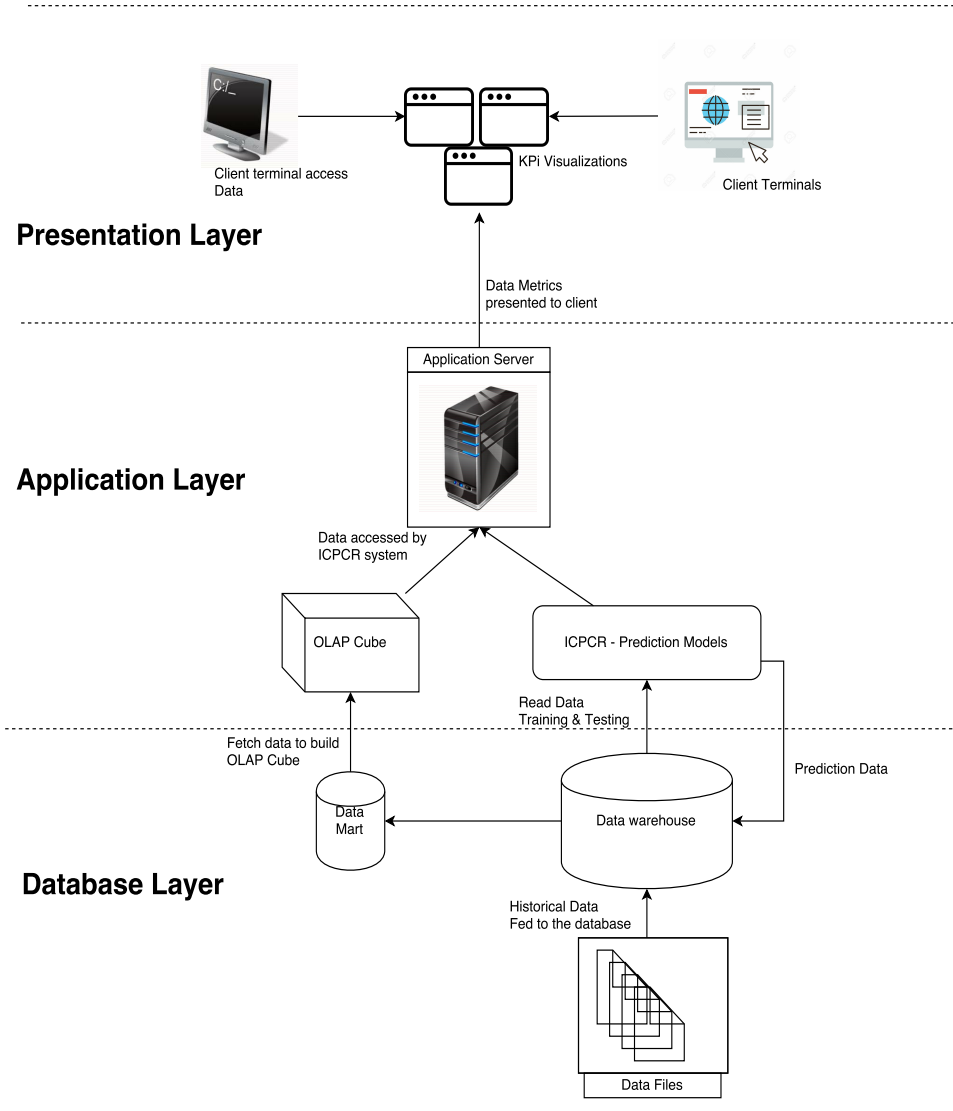
Model development & evaluation

- Model design
  - Proposal is to model 3 techniques based on Decision Tree, SVM, ANN
  - To use Machine learning libraries of either MLib, Scikit or R
- Model evaluation
  - cross validation technique to separate training and test dataset
  - Confusion matrix with scoring of Accuracy, Sensitivity and Specificity

System Development & Evaluation

4 steps to be implemented.

- Presentation Layer
  - GUI for KPI's
  - Plots of predictions
- Application Layer
  - Application processing
  - Predictive model
  - OLAP cube
- Database Layer
  - Data warehouse tables in star schema
  - Data from prediction
- System Testing
  - Unit testing



The Intelligent Churn Prediction Architecture

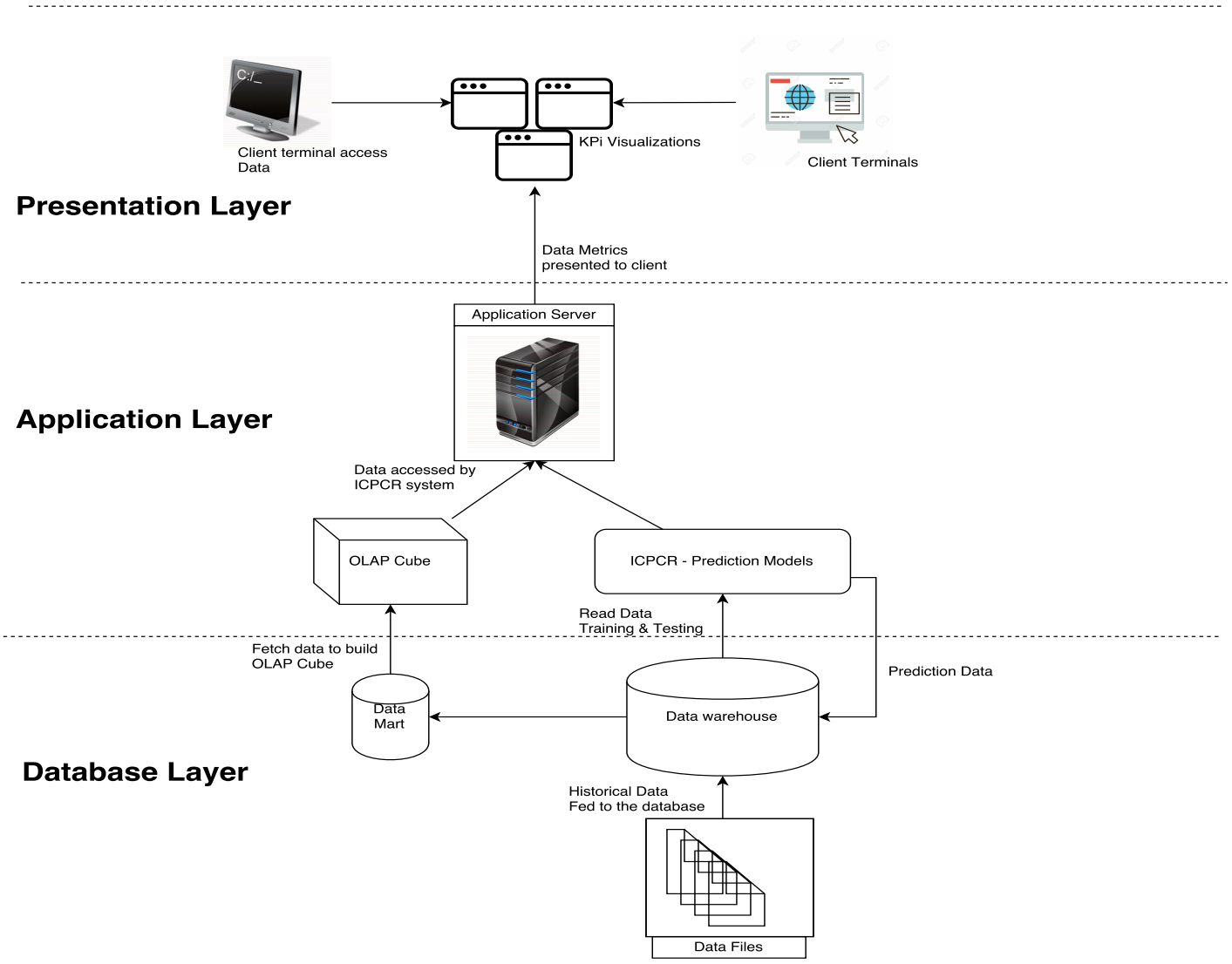


Table of Contents

Data Collection

- Churn data taken from the SGI Machine learning repository
- Data provided by Orange telecom for data analysis
- Dataset has 21 Dimensions
- Data has 5000 records of data

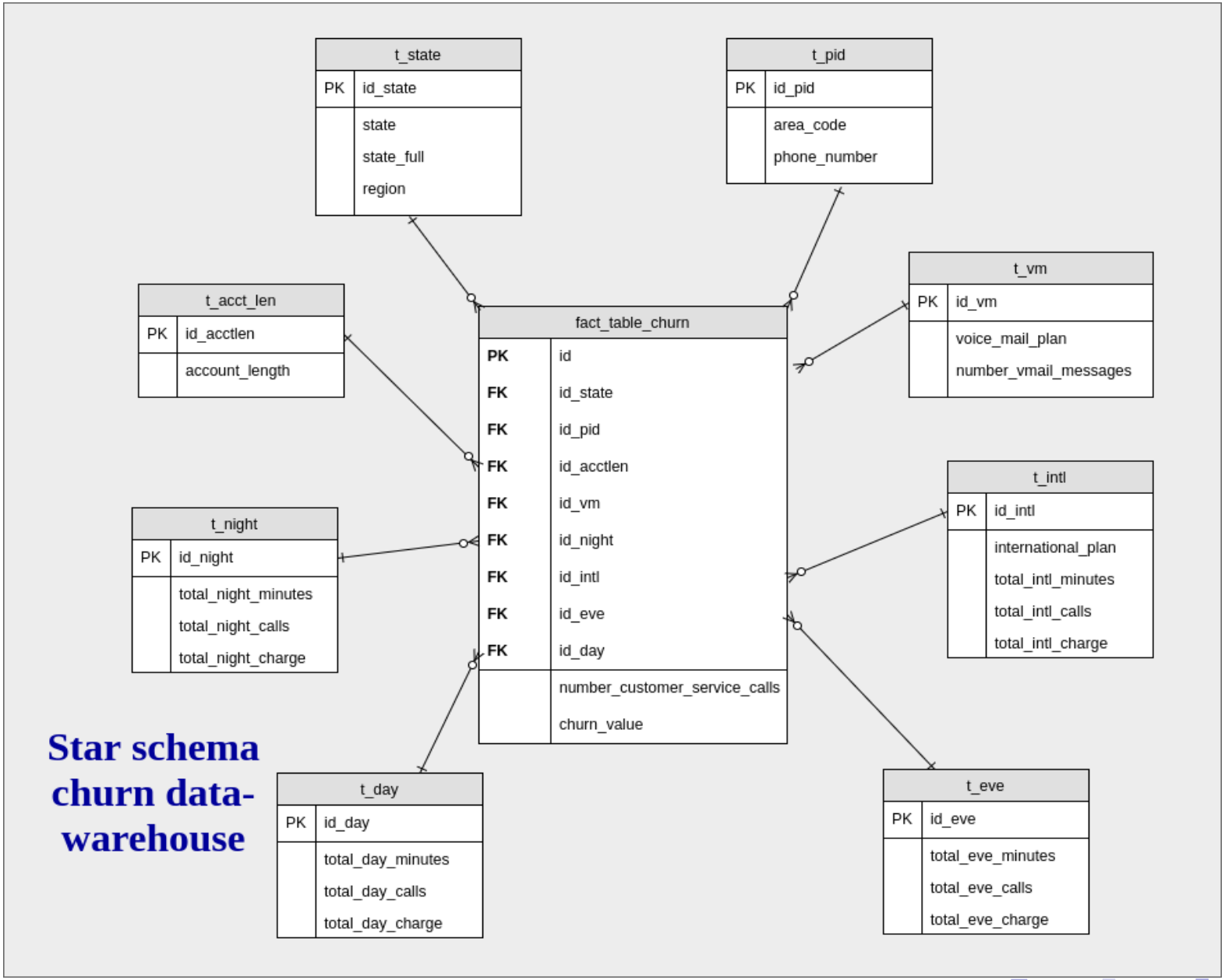
Meta-data evaluation

Serial	Name of data field	Description	Units	Type
1	State	state‘s of USA	alphabetic string	discrete
2	Account Length	months of active usage	count in months	continuous
3	Area code	area code for phone	digits	continuous
4	Phone number	phone number	digits	discrete
5	Voice mail plan	subscribed to voice mail	binary value True/False	discrete
6	Number vmail messages	number of voice-mail messages	count	continuous
7	International plan	subscribed to international plan	binary values True/False	discrete
8	Total international minutes	total number of international calls	count in minutes	continuous
9	Total international calls	total number of international calls	count	continuous
10	Total international charge	total charge of international calls	dollars	continuous
11	Total day minutes	total minutes of day calls	count in minutes	continuous
12	Total day calls	total number of day calls	count	continuous
13	Total day charge	total charge of day calls	dollars	continuous
14	Total eve minutes	total minutes of evening calls	count in minutes	continuous
15	Total eve calls	total number of evening call	count	continuous
16	Total eve charge	total charge of evening calls	dollars	continuous
17	Total night minutes	total minutes of night call	count in minutes	continuous
18	Total night calls	total number of night calls	count	continuous
19	Total night charge	total charge of night calls	dollars	continuous
20	Number customer service calls	number of calls to customer service	count	continuous
21	Churn value	if customer churned or not	binary value True/False	discrete

Data cleaning

- Important process
- Is necessary to locate unsuitable and unusable data
- Loaded into Mysql before accessed into R environment
- in R is.na functionality helps to locate and find if any row has a missing data

Data-warehouse design



Data-warehouse design

- consists of
  - 1 Fact table
  - 8 Dimension tables
- Referential integrity is maintained via Foreign key relationship

ETL process

- The etl is made up of 3 separate processes viz.
  - Extract - Data is extracted into MySQL tables and the taken into R environment for processing
  - Transform - Data needs to be converted from dataframe format to matrix format, or some categorical string variables need to be converted to numeric values
  - Load - Loading the resultant data back to csv files and MySQL tables to preserve data from corruption

Table of Contents

Prediction models

8 classification machine learning algorithms were trained and tested.

- Decision tree from rpart, ctree, C5.0, boosted ctree,
- Random forest,
- Naive Bayes,
- Support Vector Machine,
- Neural networks

Regression and Paritioning tree

Table: Confusion matrix for rpart decision tree

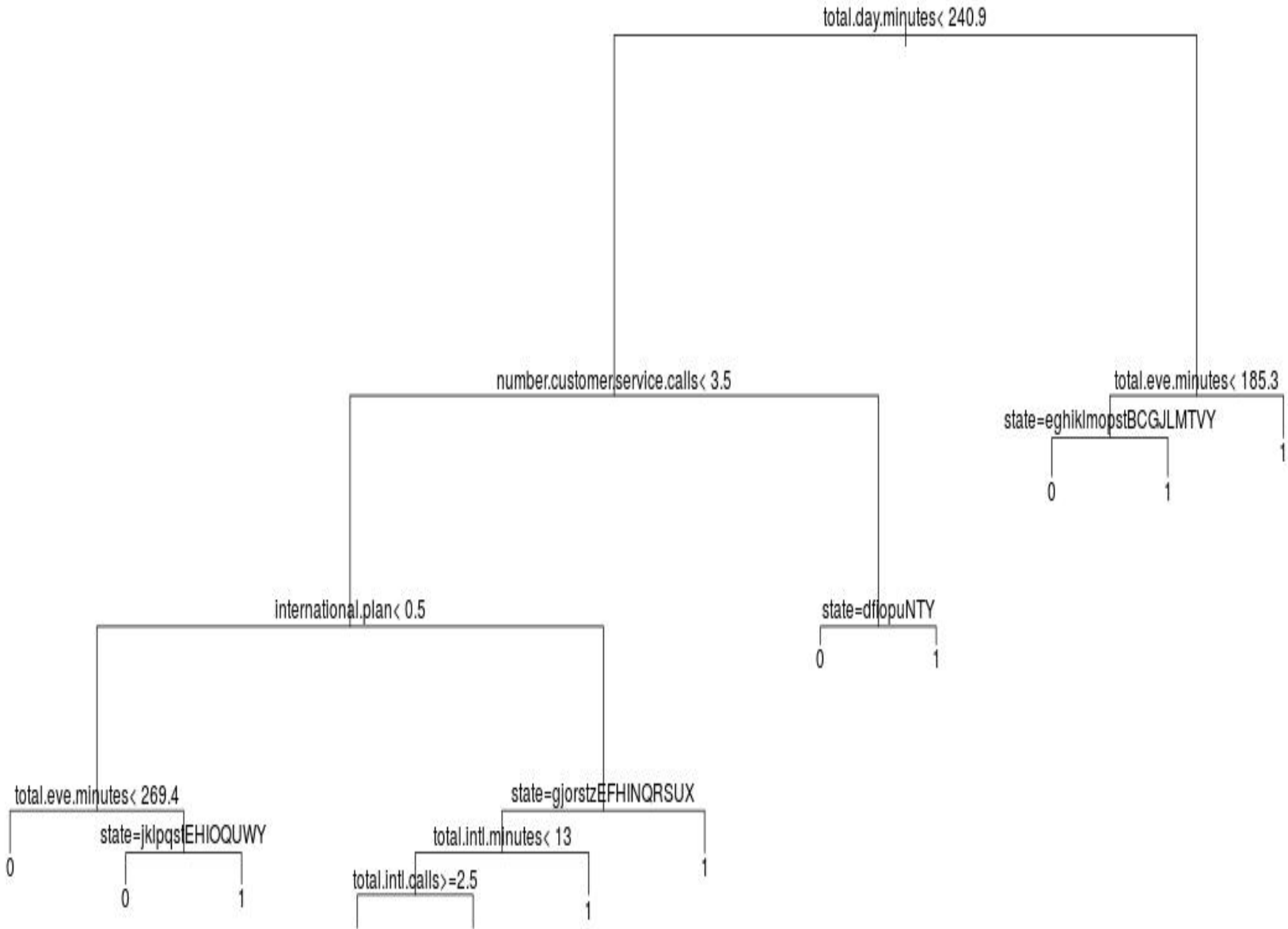
	Actual	
Prediction	False 0	True 1
False 0	194	36
True 1	29	165

Rpart decison tree performance indicators

Table: Accuracy Sensitivity and Specificity

Accuracy	:	0.8466
Sensitivity	:	0.8699552
Specificity	:	0.8208955

Regression and Paritioning tree plot



Conditional inference tree

Table: Confusion matrix for ctree decision tree

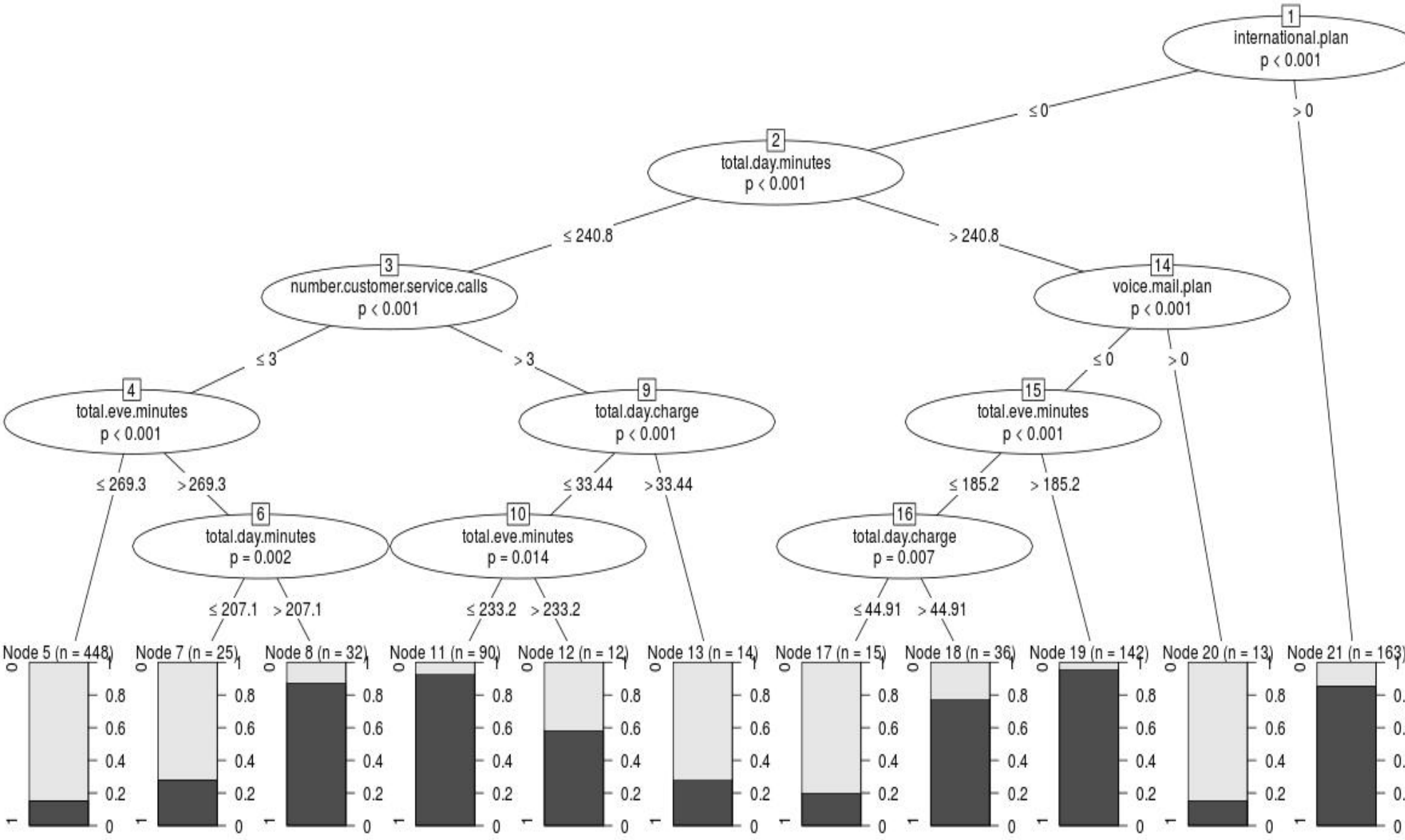
	Actual	
Prediction	False 0	True 1
False 0	206	26
True 1	17	175

Ctree decison tree performance indicators

Table: Accuracy Sensitivity and Specificity

Accuracy	:	0.8985849
Sensitivity	:	0.9237668
Specificity	:	0.8706468

Conditional inference tree plot



Random forests

Table: Confusion matrix for random forests

	Actual	
Prediction	False 0	True 1
False 0	216	74
True 1	7	127

Random forest performance indicators

Table: Accuracy Sensitivity and Specificity

Accuracy	:	0.8089623
Sensitivity	:	0.9686099
Specificity	:	0.6318408

C 5.0 decision tree

Table: Confusion matrix for C5.0 tree

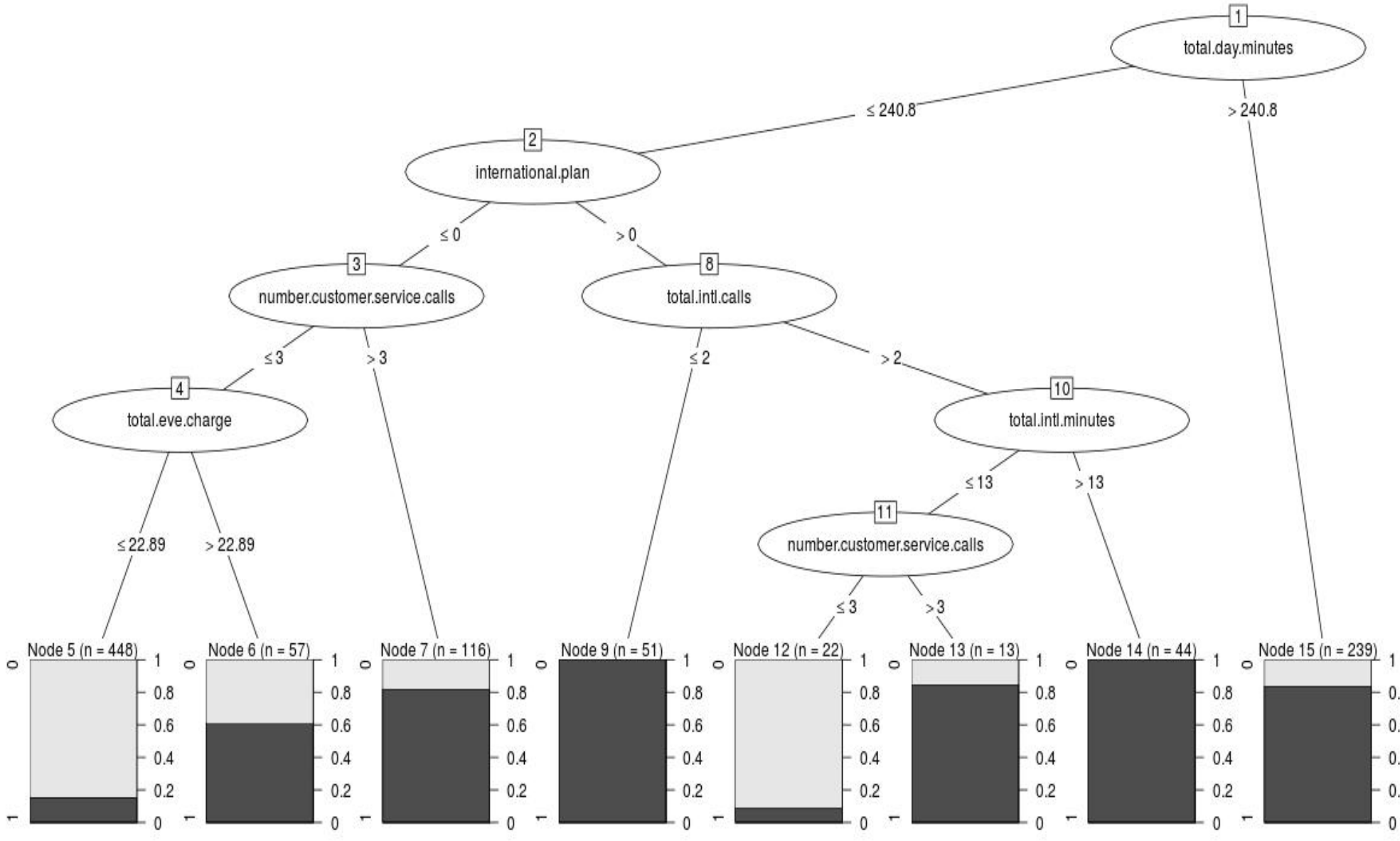
	Actual	
Prediction	False 0	True 1
False 0	182	24
True 1	41	177

C5.0's performance indicators

Table: Accuracy Sensitivity and Specificity

Accuracy	:	0.8466981
Sensitivity	:	0.8161435
Specificity	:	0.8805970

C 5.0 decision tree plot



Boosted C5.0 decision tree

Table: Confusion matrix for Boosted C5.0 tree

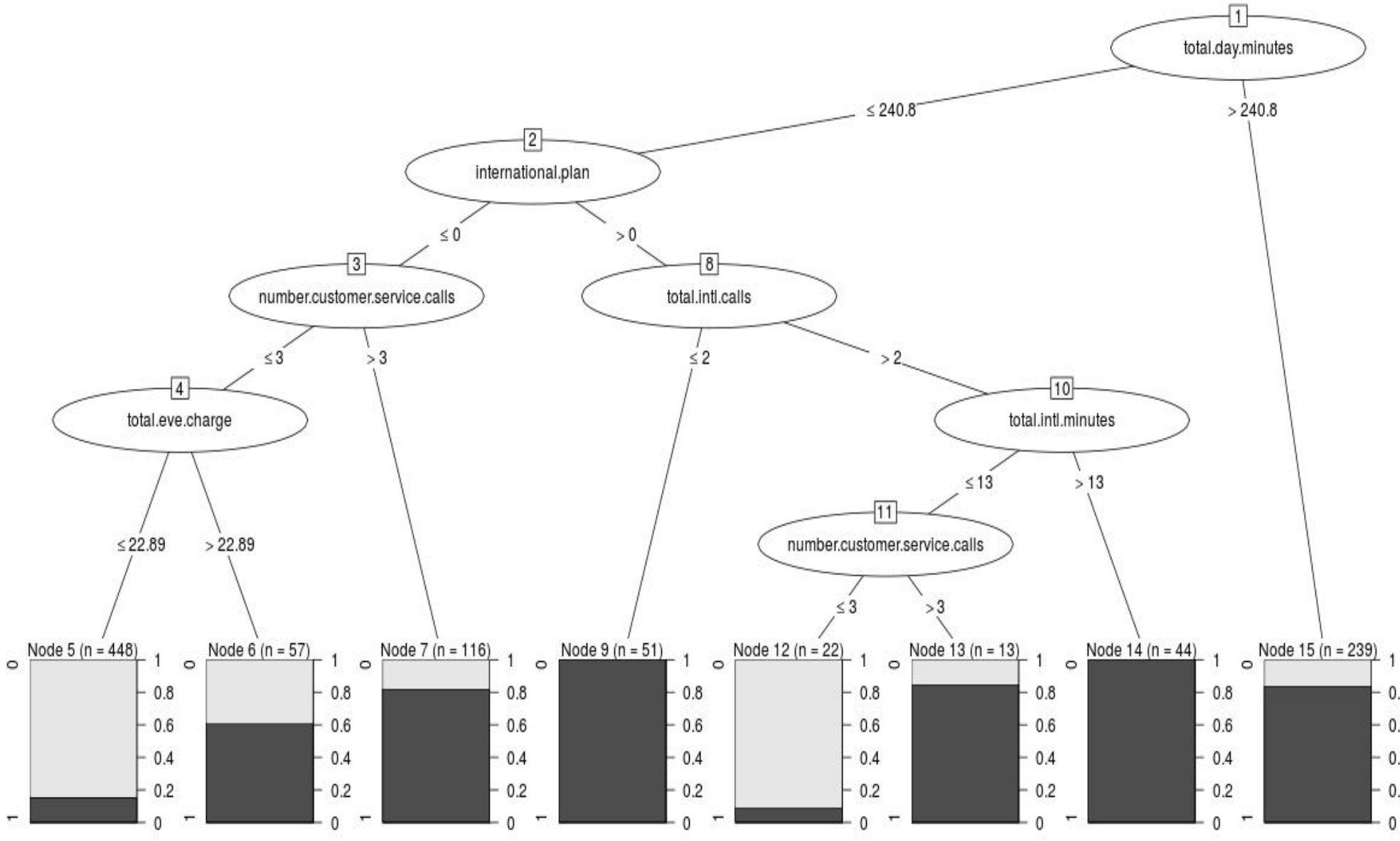
	Actual	
Prediction	False 0	True 1
False 0	200	22
True 1	23	179

Boosted C5.0's performance indicators

Table: Accuracy Sensitivity and Specificity

Accuracy	:	0.8938679
Sensitivity	:	0.8968610
Specificity	:	0.8905473

Boosted C 5.0 decision tree plot



Naive Bayes classifier

Table: Confusion matrix for Naive Bayes

	Actual	
Prediction	False 0	True 1
False 0	177	41
True 1	40	160

Naive Bayes performance indicators

Table: Accuracy Sensitivity and Specificity

Accuracy	:	0.7948113
Sensitivity	:	0.7937220
Specificity	:	0.7960199

SVM

Table: Confusion matrix for SVM

	Actual	
Prediction	False 0	True 1
False 0	171	29
True 1	52	172

SVM performance indicators

Table: Accuracy Sensitivity and Specificity

Accuracy	:	0.8089623
Sensitivity	:	0.7668161
Specificity	:	0.8557214

Neural network

A neural network with 4 and 3 nodes in 1st and 2nd layer respectively.

Table: Confusion matrix for NN

	Actual	
Prediction	False 0	True 1
False 0	168	29
True 1	55	172

NN performance indicators

Table: Accuracy Sensitivity and Specificity

Accuracy	:	0.8655
Sensitivity	:	0.7668
Specificity	:	0.8557

Neural network plot

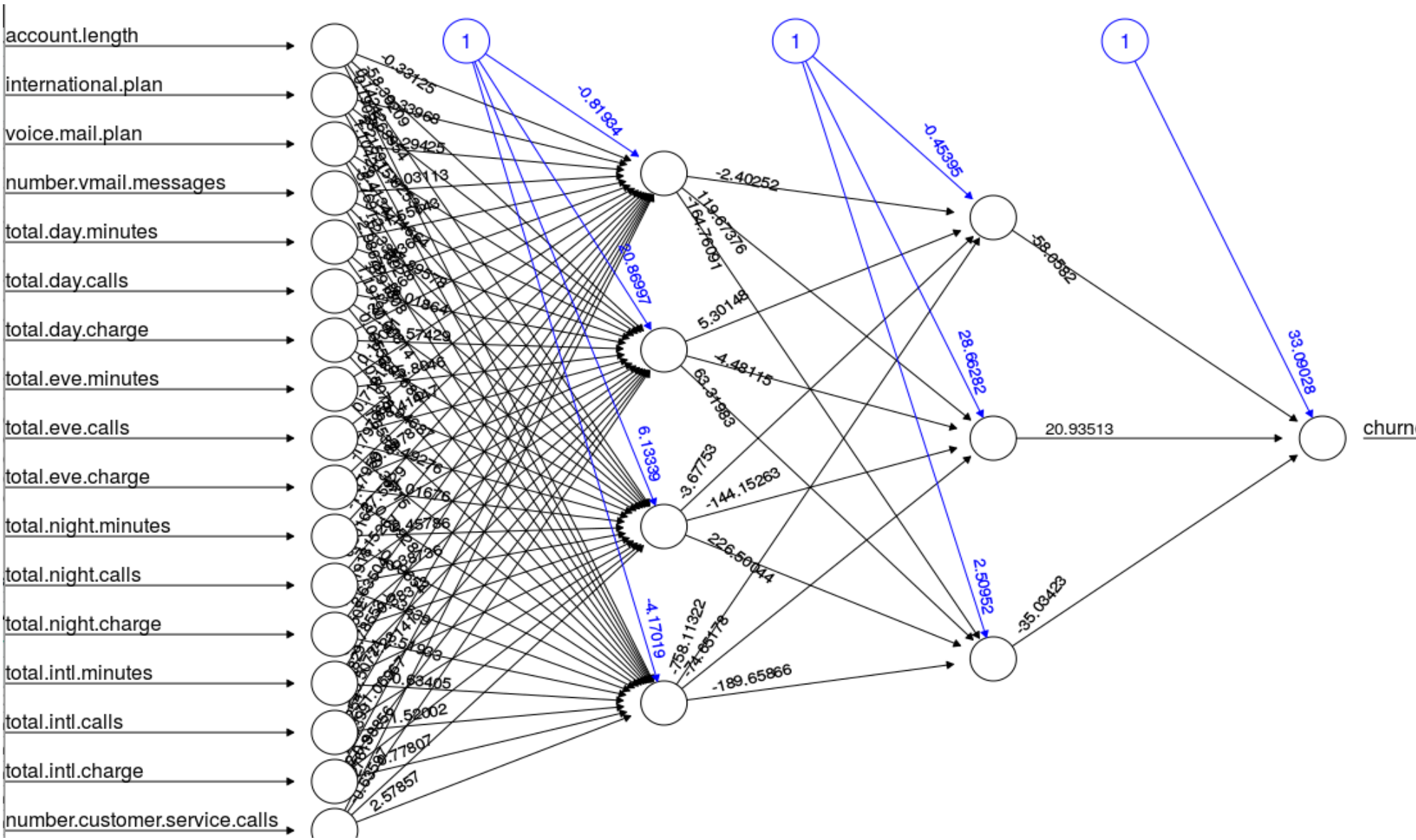


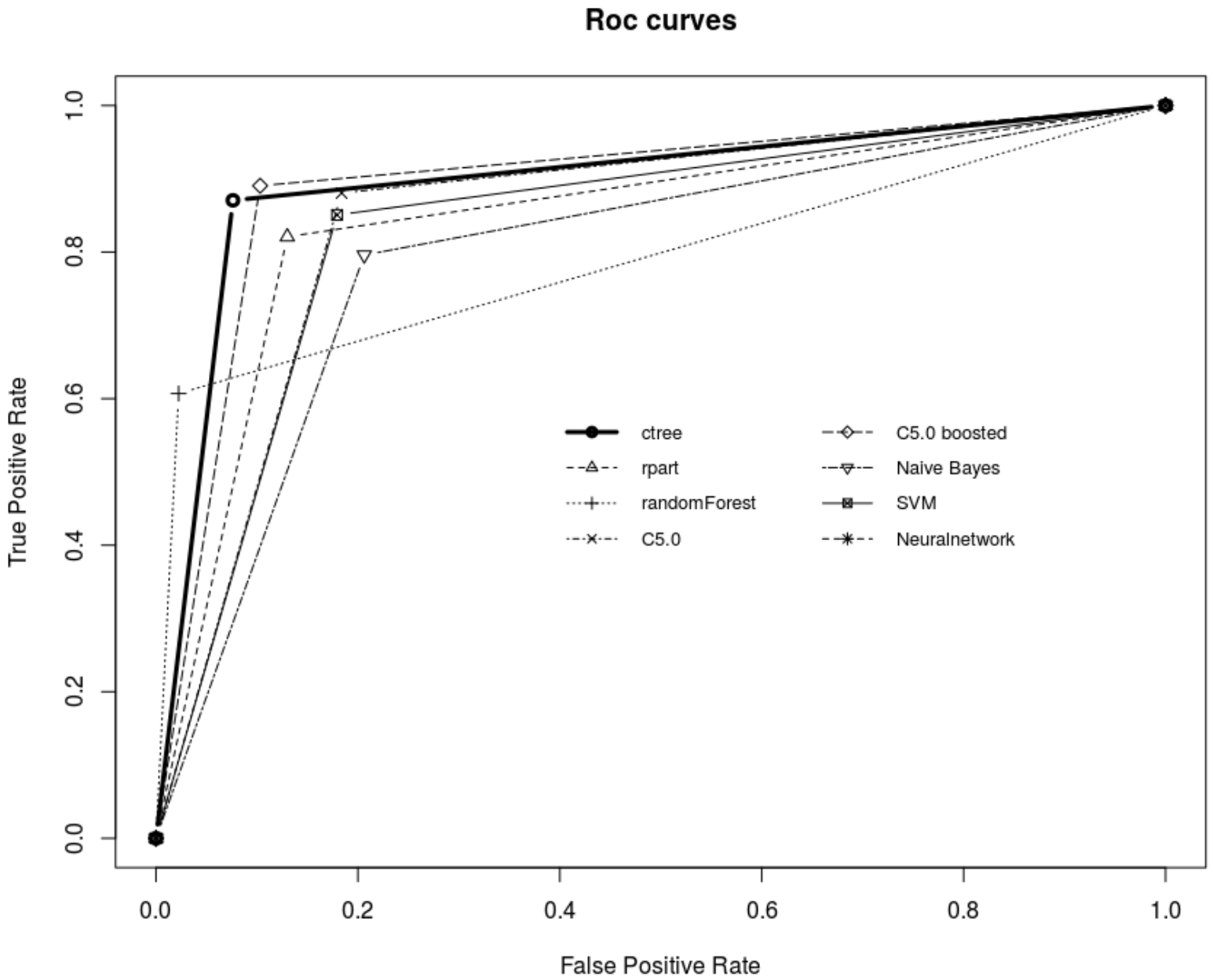
Table of Contents

Comparision of metrics between classification models

Table: Performance comparisons for prediction models

Model	Accuracy	Sensitivity	Specificity	Pos Pred Ra
Tree - Rpart	0.8467	0.8700	0.8209	0.8435
Tree - Ctree	0.8986	0.9238	0.8706	0.8879
Tree - Ctree Boosted	0.8774	0.8969	0.8557	0.8734
Random Forest	0.7948	0.9641	0.6070	0.7313
C5.0	0.8467	0.8161	0.8806	0.8835
Naive Bayes	0.7948	0.7937	0.7960	0.8119
Support Vector Machines	0.8089	0.8286	0.8557	0.8550
Neural Network	0.8655	0.7668	0.8557	0.8429

Roc comparision



Decision Rules

```
[1] root
|   [2] international.plan <= 0
|   |   [3] total.day.minutes <= 240.8
|   |   |   [4] number.customer.service.calls <= 3
|   |   |   |   [5] total.eve.minutes <= 269.3: 0 (n = 448, err = 15.1786%)
|   |   |   |   [6] total.eve.minutes > 269.3
|   |   |   |   |   [7] total.day.minutes <= 207.1: 0 (n = 25, err = 28.0000%)
|   |   |   |   |   [8] total.day.minutes > 207.1: 1 (n = 32, err = 12.5000%)
|   |   |   [9] number.customer.service.calls > 3
|   |   |   |   [10] total.day.charge <= 33.44
|   |   |   |   |   [11] total.eve.minutes <= 233.2: 1 (n = 90, err = 6.6667%)
|   |   |   |   |   [12] total.eve.minutes > 233.2: 1 (n = 12, err = 41.6667%)
|   |   |   |   [13] total.day.charge > 33.44: 0 (n = 14, err = 28.5714%)
|   |   [14] total.day.minutes > 240.8
|   |   |   [15] voice.mail.plan <= 0
|   |   |   |   [16] total.eve.minutes <= 185.2
|   |   |   |   |   [17] total.day.charge <= 44.91: 0 (n = 15, err = 20.0000%)
|   |   |   |   |   [18] total.day.charge > 44.91: 1 (n = 36, err = 22.2222%)
|   |   |   |   [19] total.eve.minutes > 185.2: 1 (n = 142, err = 4.2254%)
|   |   |   [20] voice.mail.plan > 0: 0 (n = 13, err = 15.3846%)
|   [21] international.plan > 0: 1 (n = 163, err = 14.7239%)
```

Table of Contents

Intelligent churn prediction and customer retention system - ICPCR

- The system is designed as a Web app with RShiny
- Data mining models were sourced from libraries and developed in R programming
- All the pages and plots are rendered by the R scripts
- Databse used is MySQL

Figure: Dashboard - Data explorer

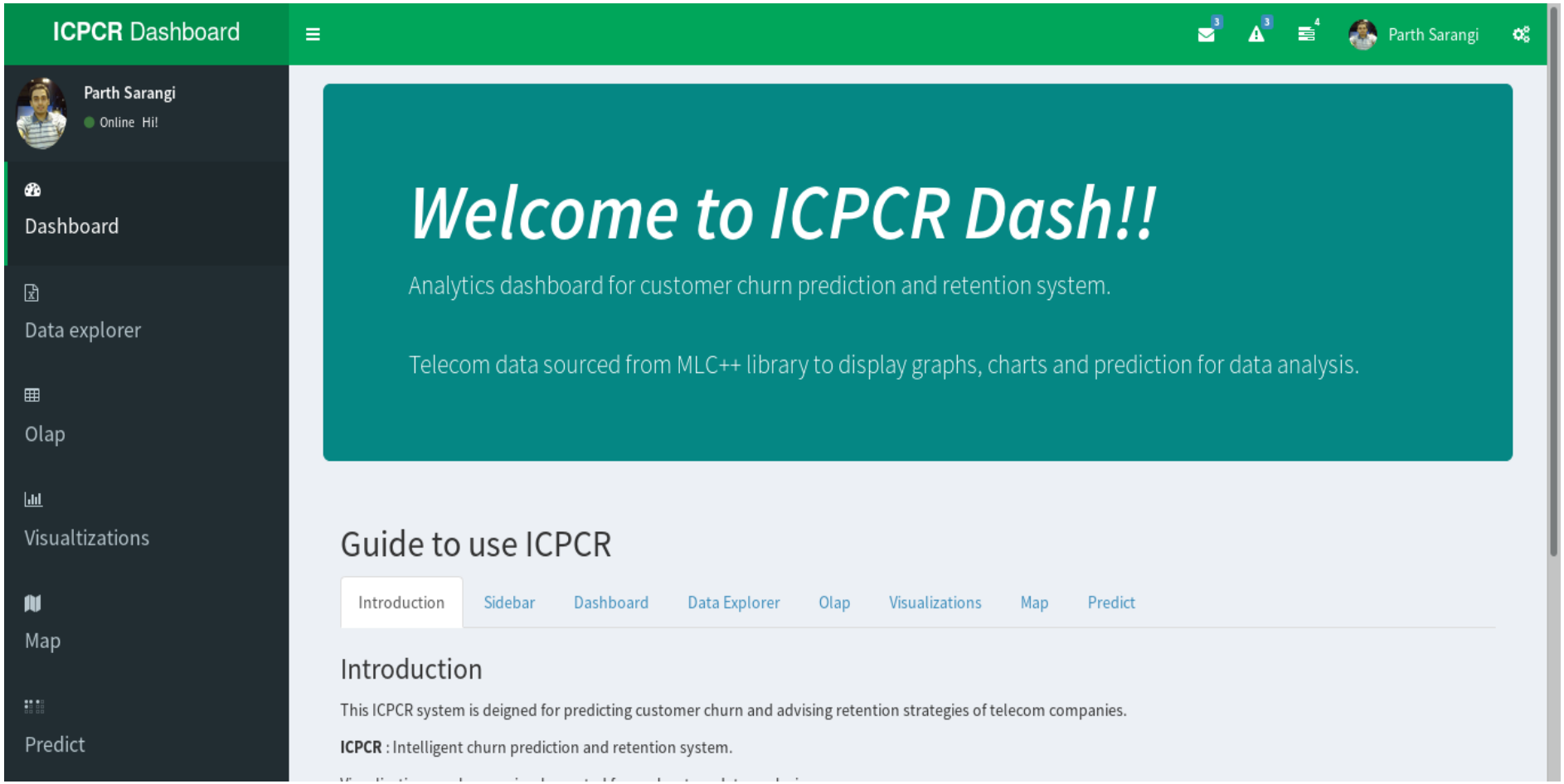


Figure: Dashboard - Key performance indicators

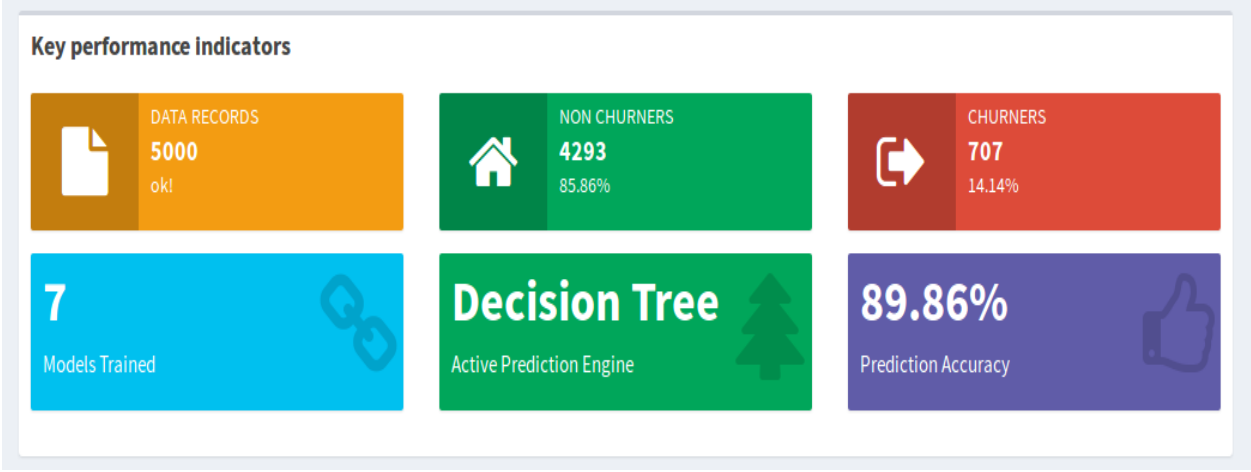
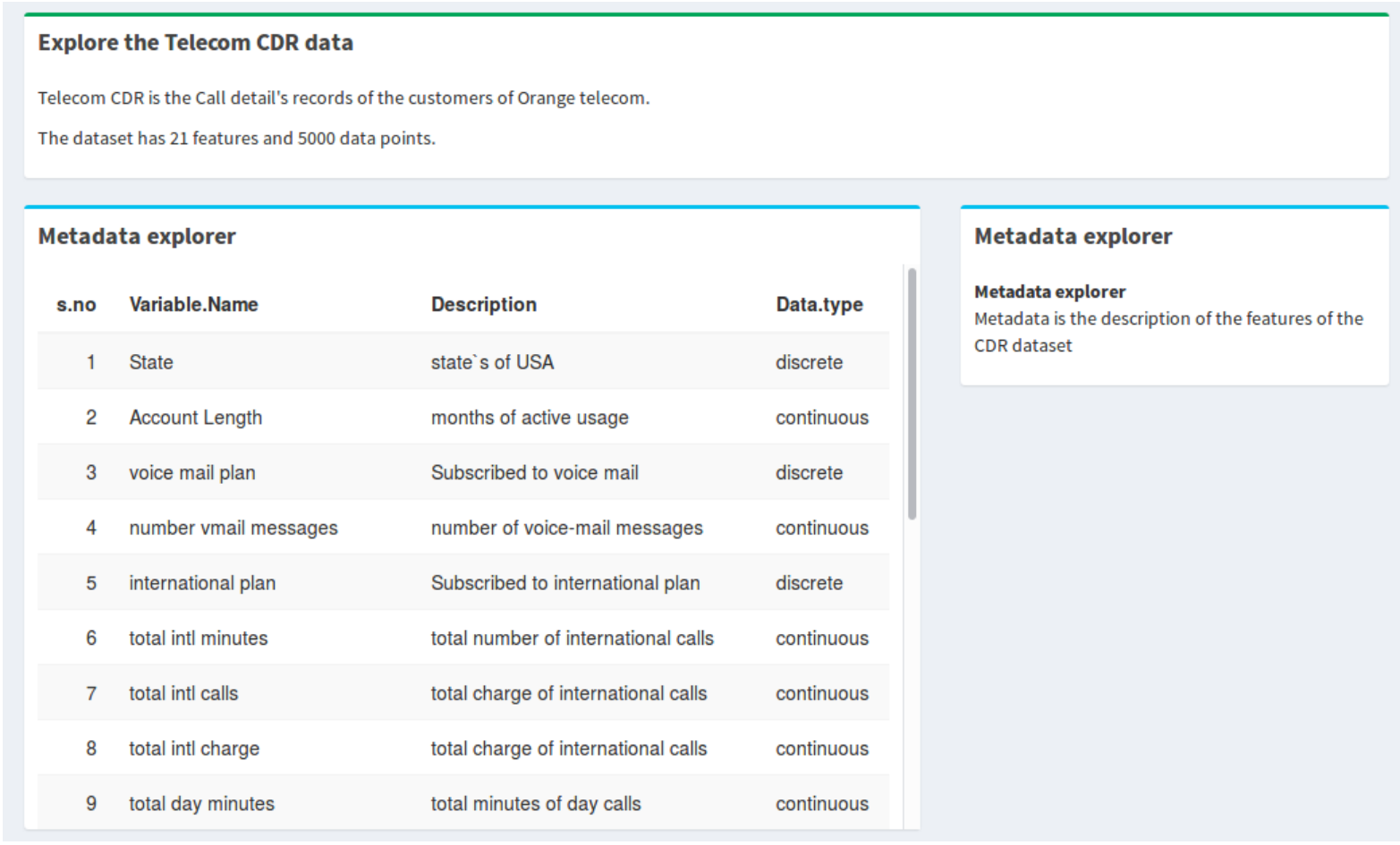


Figure: Dashboard - Field analysis



Explore the Telecom CDR data

Telecom CDR is the Call detail's records of the customers of Orange telecom.

The dataset has 21 features and 5000 data points.

Metadata explorer

s.no	Variable.Name	Description	Data.type
1	State	state`s of USA	discrete
2	Account Length	months of active usage	continuous
3	voice mail plan	Subscribed to voice mail	discrete
4	number vmail messages	number of voice-mail messages	continuous
5	international plan	Subscribed to international plan	discrete
6	total intl minutes	total number of international calls	continuous
7	total intl calls	total charge of international calls	continuous
8	total intl charge	total charge of international calls	continuous
9	total day minutes	total minutes of day calls	continuous

Metadata explorer

Metadata explorer

Metadata is the description of the features of the CDR dataset

Figure: Dashboard - Bar plot of Total, Churner and non-Churner customer counts region wise

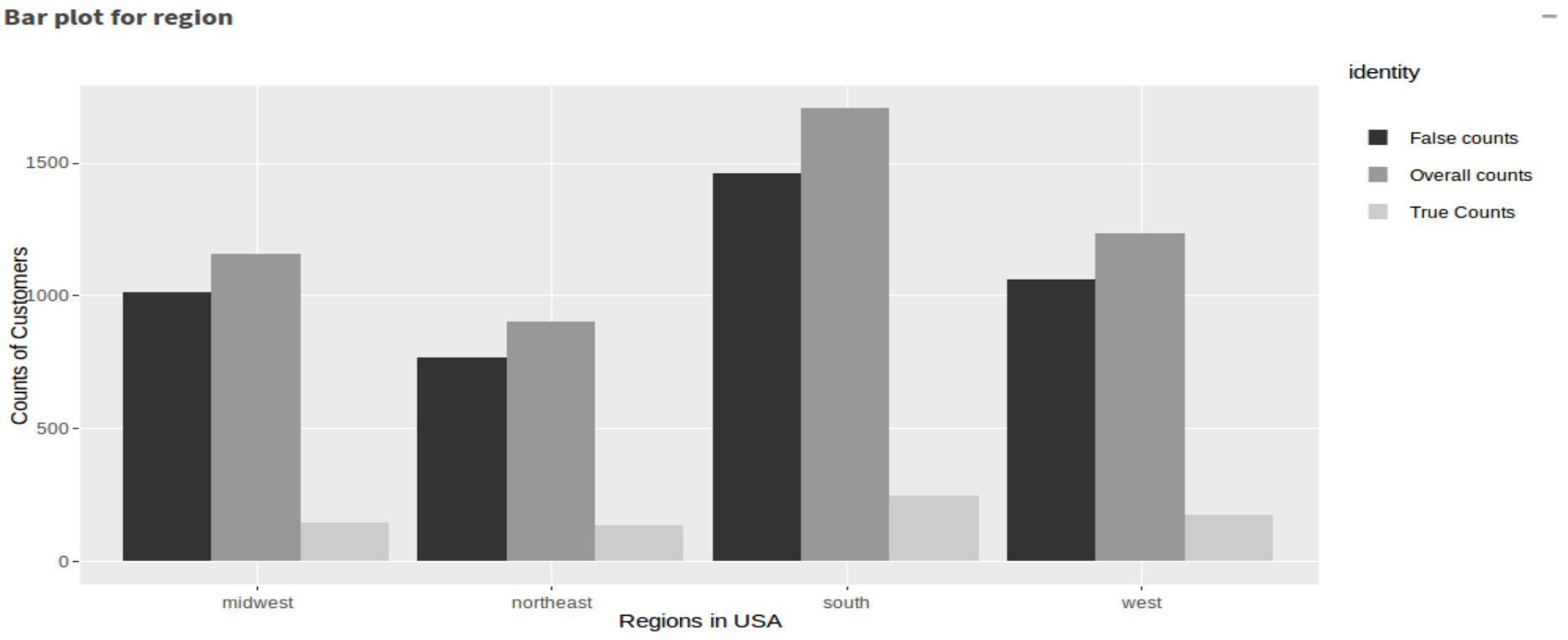
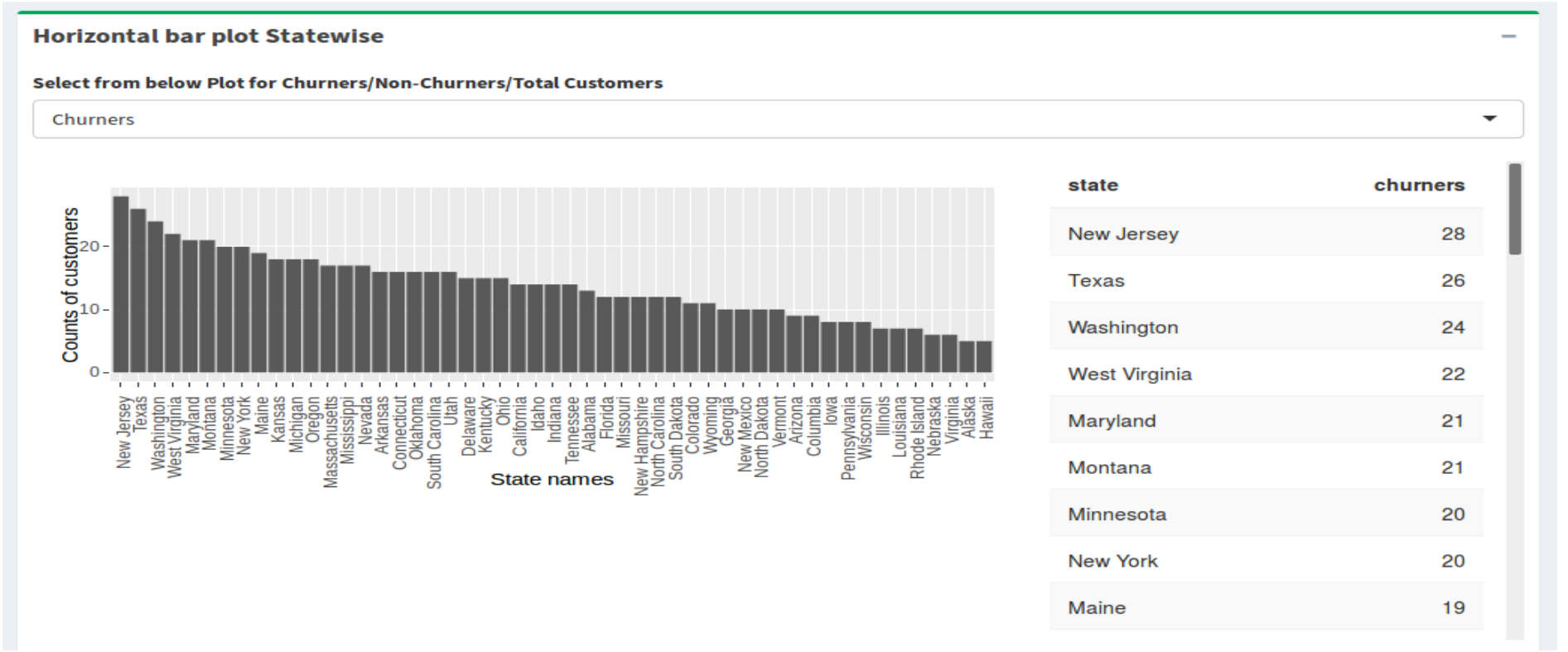


Figure: Dashboard - Bar plot of Churner customers to Non-churners State wise



state	churners
New Jersey	28
Texas	26
Washington	24
West Virginia	22
Maryland	21
Montana	21
Minnesota	20
New York	20
Maine	19

Figure: Dashboard - Map to visually show the density of Churners and Non-churners state wise - Part 1

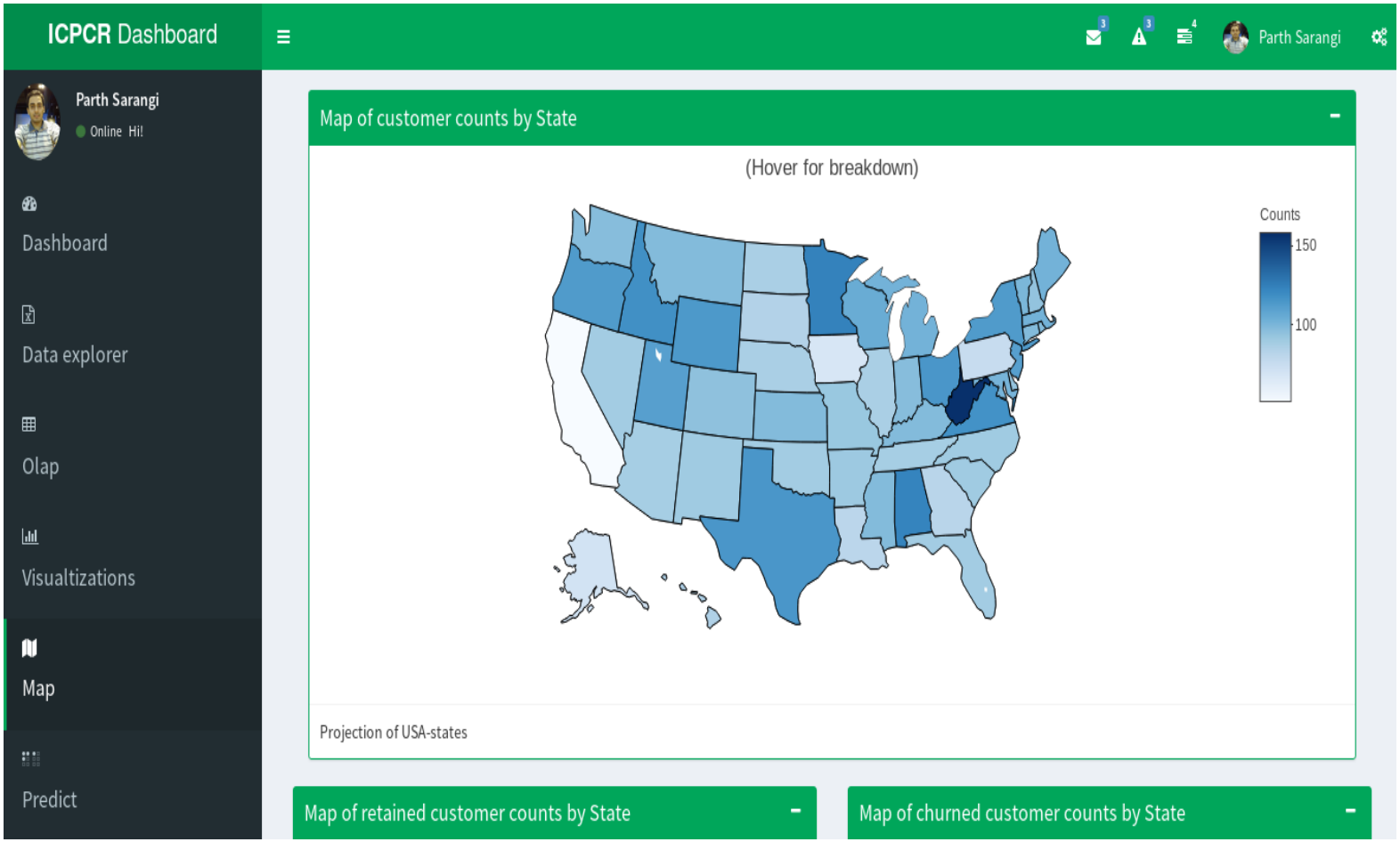


Figure: Dashboard - Map to visually show the density of Churners and Non-churners state wise - Part 2

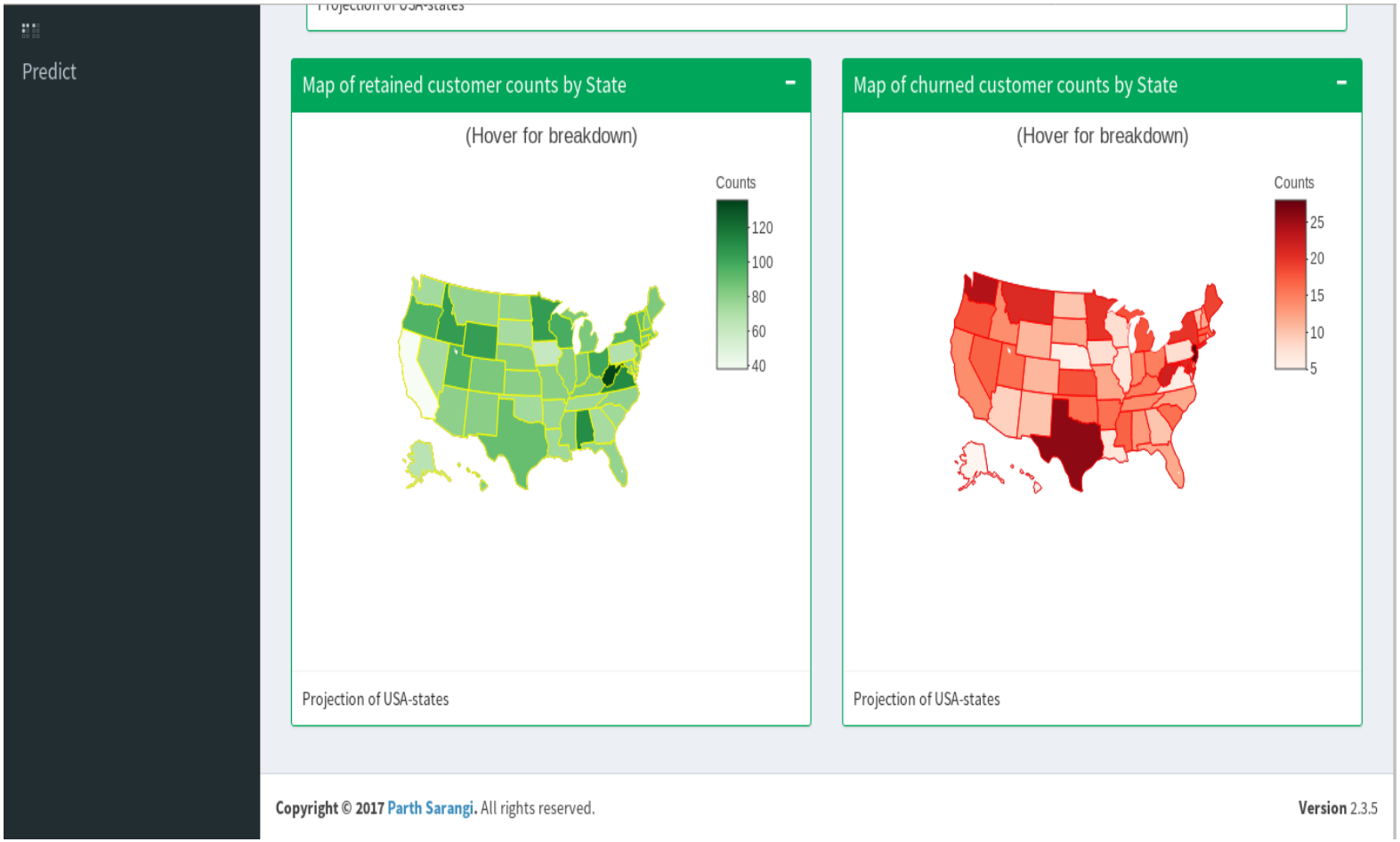


Figure: Dashboard - Upload of file for prediction

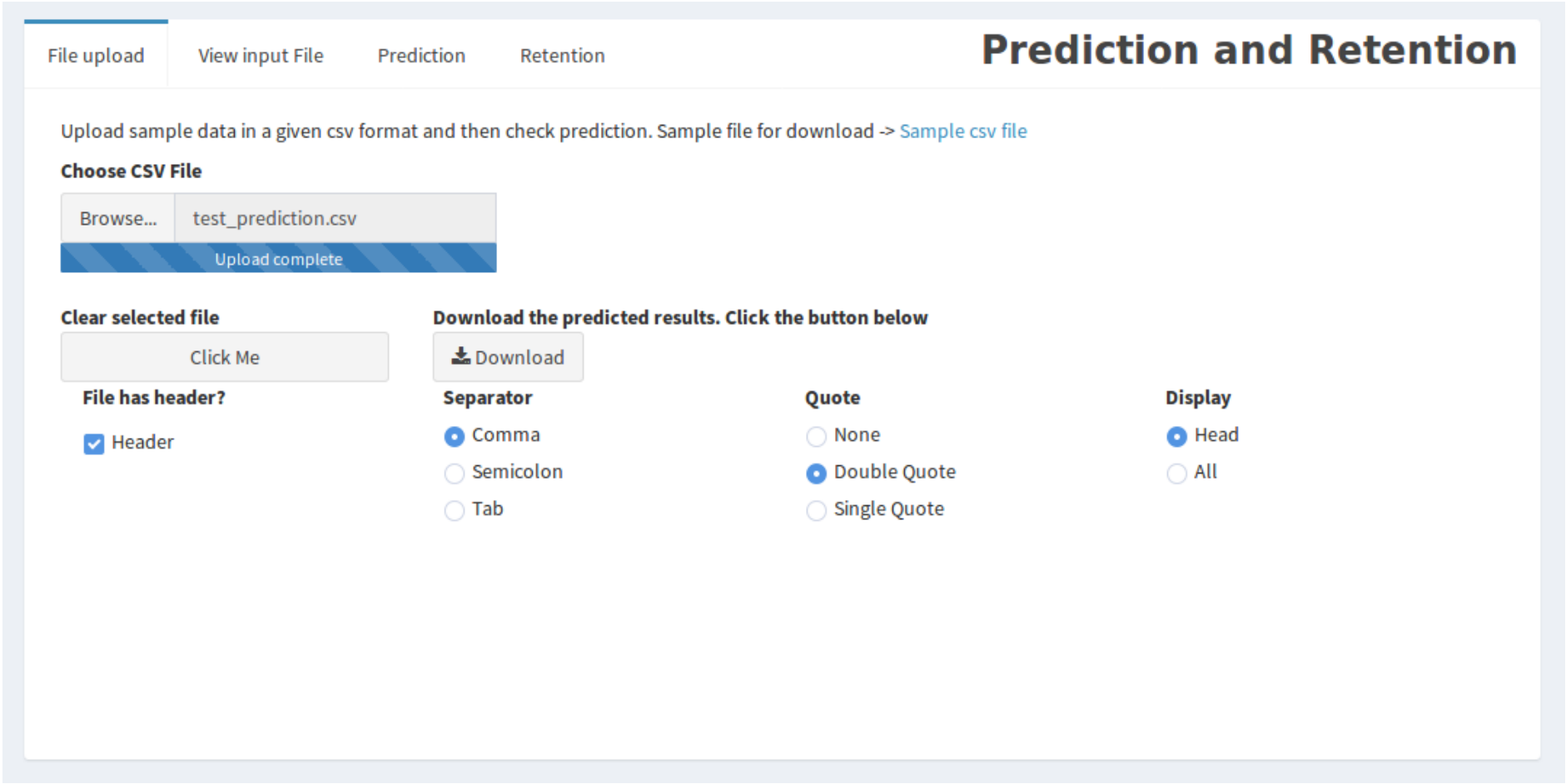


Figure: Dashboard - View uploaded file contents

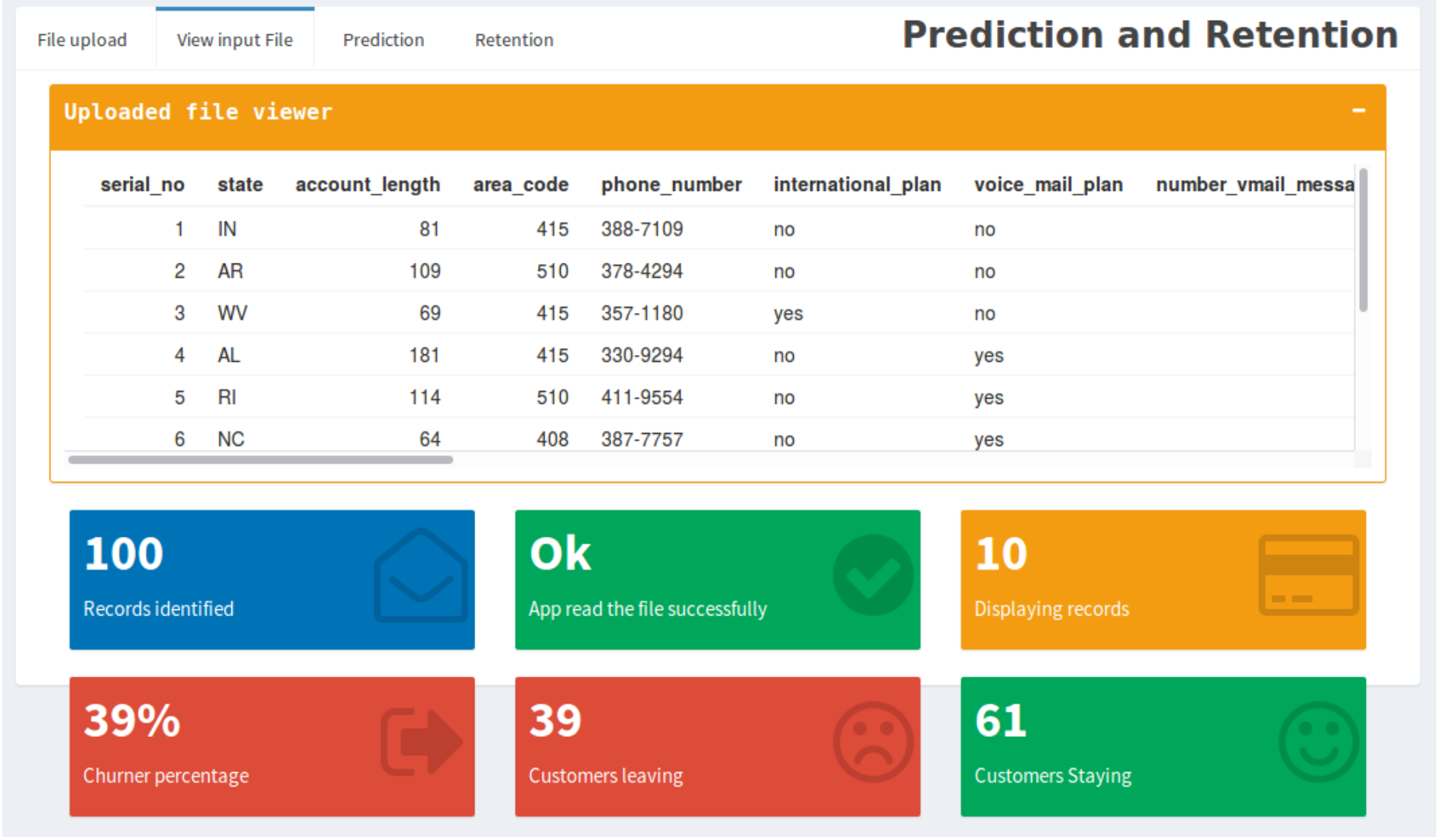
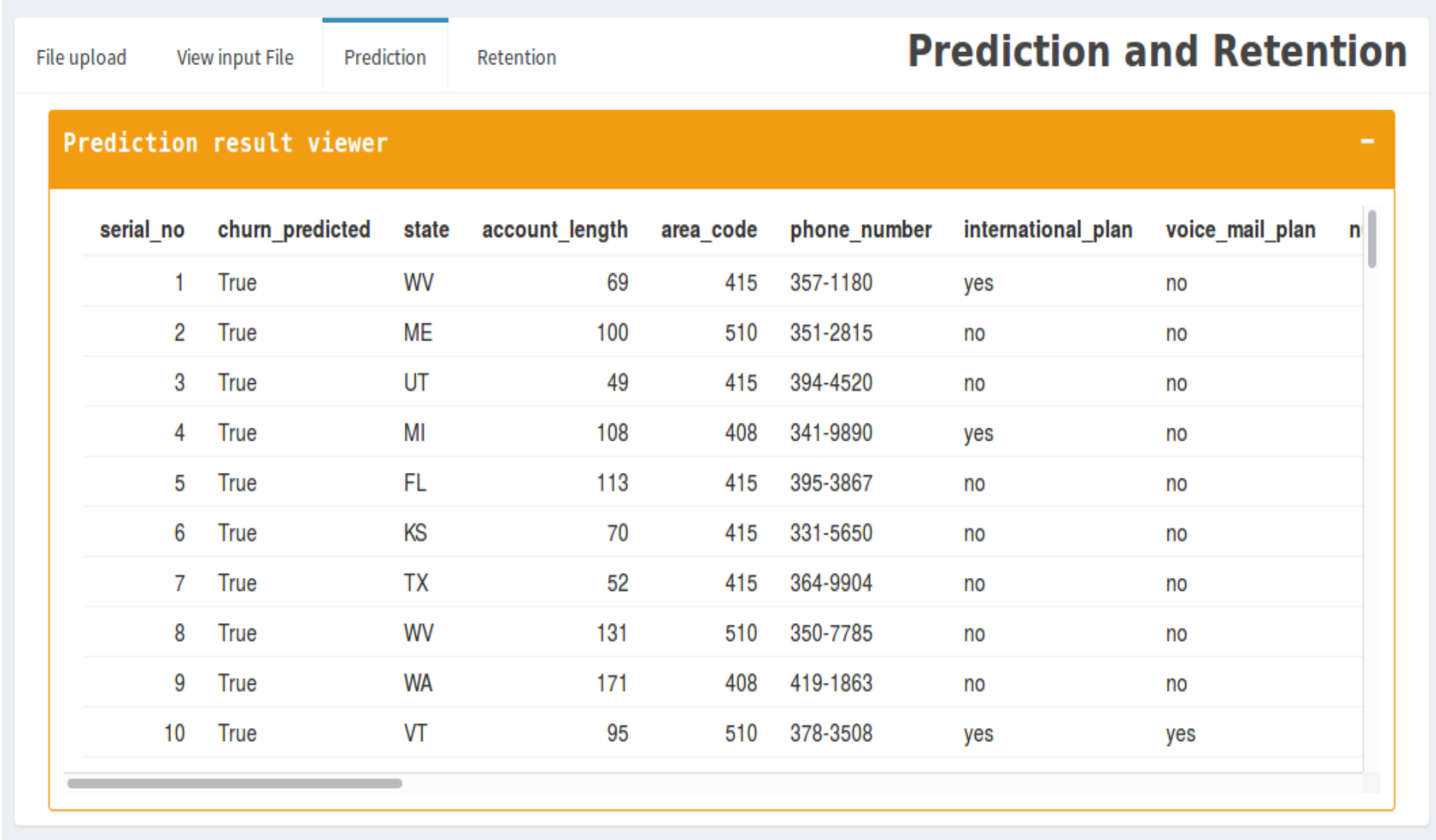
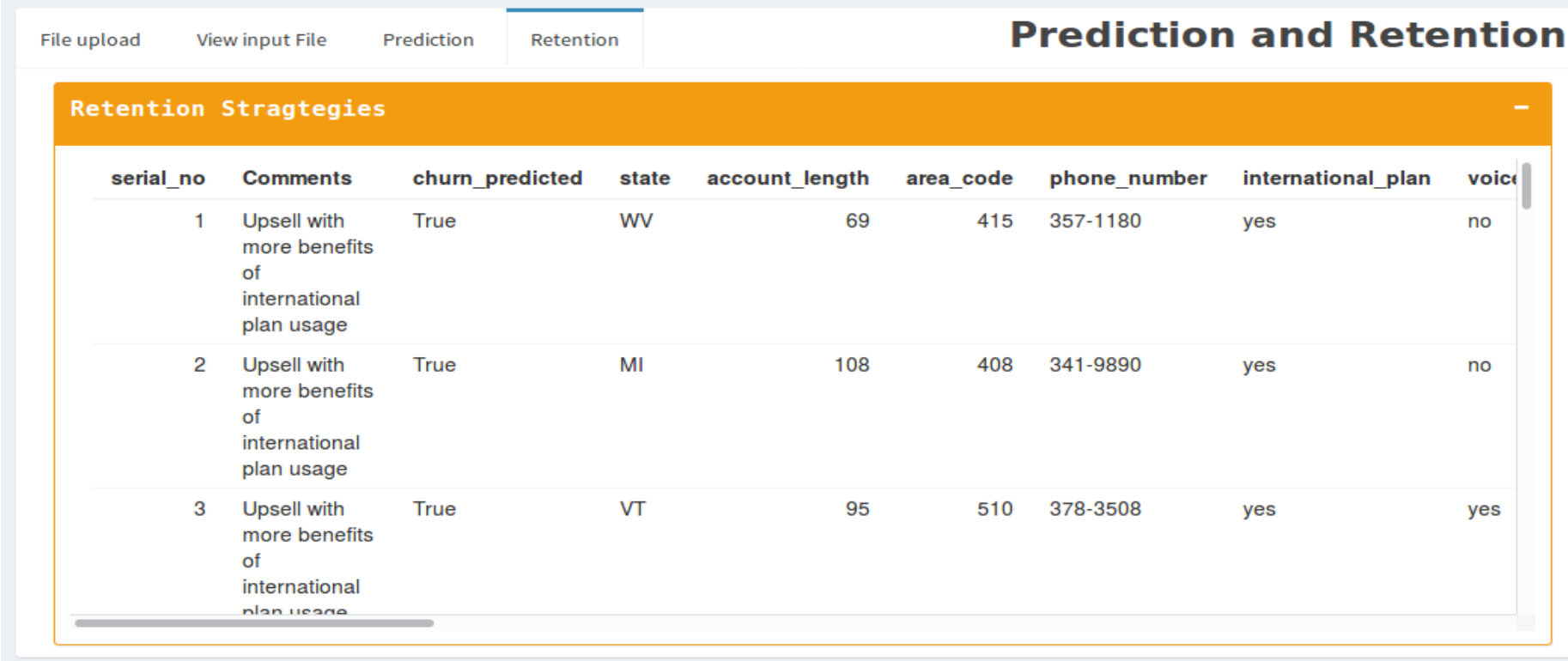


Figure: Dashboard - View prediction results



serial_no	churn_predicted	state	account_length	area_code	phone_number	international_plan	voice_mail_plan	n
1	True	WV	69	415	357-1180	yes	no	
2	True	ME	100	510	351-2815	no	no	
3	True	UT	49	415	394-4520	no	no	
4	True	MI	108	408	341-9890	yes	no	
5	True	FL	113	415	395-3867	no	no	
6	True	KS	70	415	331-5650	no	no	
7	True	TX	52	415	364-9904	no	no	
8	True	WV	131	510	350-7785	no	no	
9	True	WA	171	408	419-1863	no	no	
10	True	VT	95	510	378-3508	yes	yes	

Figure: Dashboard - View retention strategies



serial_no	Comments	churn_predicted	state	account_length	area_code	phone_number	international_plan	voic
1	Upsell with more benefits of international plan usage	True	WV	69	415	357-1180	yes	no
2	Upsell with more benefits of international plan usage	True	MI	108	408	341-9890	yes	no
3	Upsell with more benefits of international plan usage	True	VT	95	510	378-3508	yes	yes

Table of Contents

Conclusion

- Used open source dataset from SGI library with 5000 records and 21 features
- Irrelevant features were removed with only 18 features for building models
- Feature selection technique to build suitable data mining model
- Trained, tested and compared 8 classification models
- Conditional inference decision tree seems most suitable with highest accuracy of 89%.
- Customer retention is implemented with upselling strategies
- Upselling strategies are developed from decision rules of the decision tree

Recommendation

- Data mining is huge field and research opportunities are plenty.
- Model comparison with feature selection techniques can be explored.
- Data mining from other domains such as medicine, natural calamities, consumption analysis
- We also suggest looking into Natural language processing
- Also suggest to research predictive analytics in Blockchains