Telecom Churn Analysis – A Case Study

Cognizant

Introduction

- Global Top Trends in technology and computing includes mobile technology

- Landscape of Telecom Industry has changed
  - ✓ Large Number of Private Service Providers have evolved
  - ✓ To Survive in current Scenario new innovative business models are a must

- Churn is huge factor in Telecom Industry

- Major initiators of churn include
  - ✓ Quality of service
  - ✓ Tariffs
  - ✓ Dissatisfaction in post sales service etc.

- Interesting facts surrounding churn
  - ✓ Annual churn rate is estimated to be 25-30% in Europe
  - ✓ Acquiring new customers is costlier than retaining them

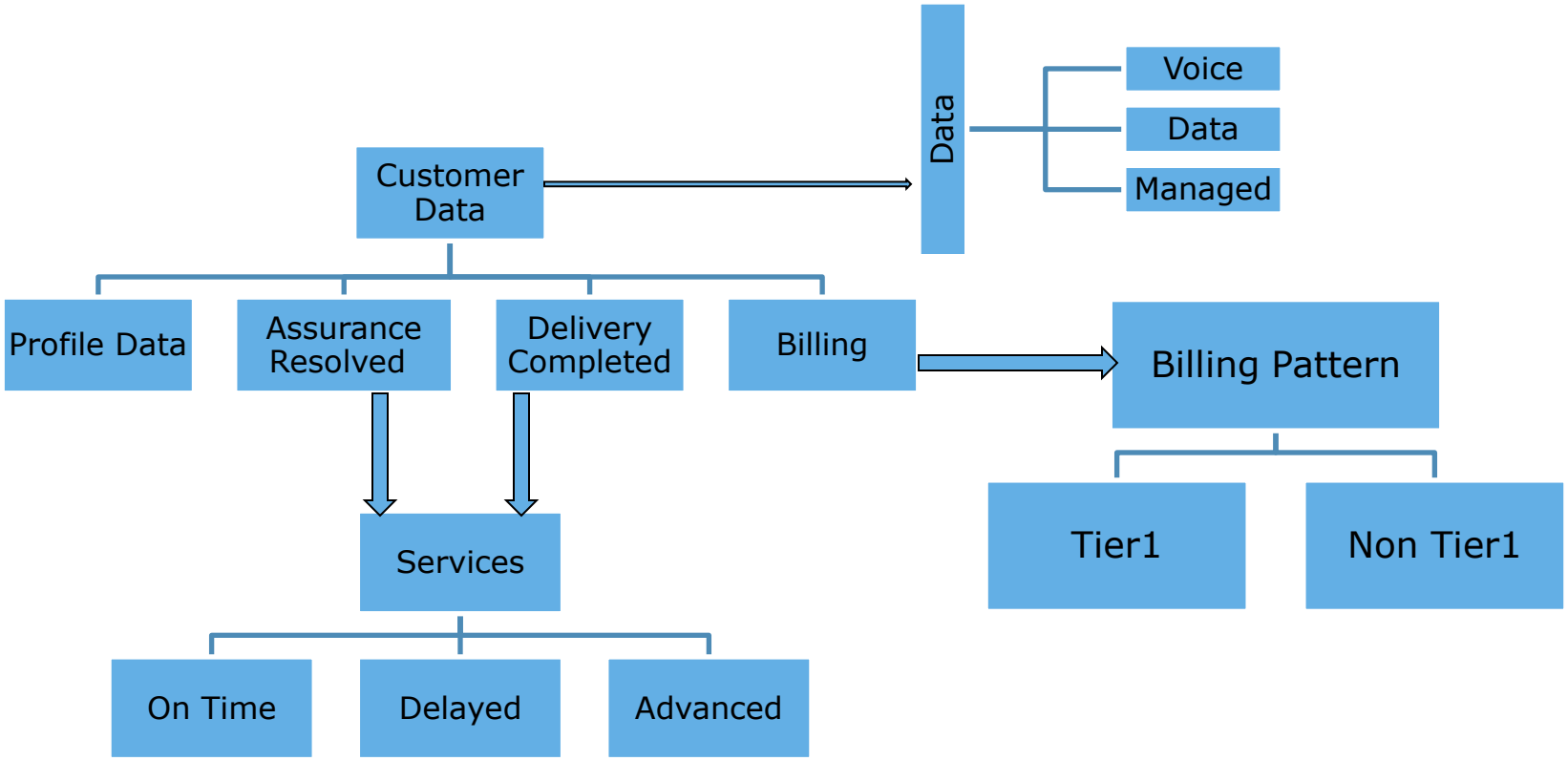
Cognizant

Objective

Objective of the current study was to predict churn and identify the key drivers of churn in each business division using simulated customer data sets.

Cognizant

Telecom Data - Classification



Cognizant

Analysis Data

Records used in the study

Table Name	Initial Records – Before Cleaning	Records – After Cleaning
Assurance Resolved	176743	176072
Delivery Completed	177795	136695
Billing	17,64,883	1764883

Sample Key Attributes of Assurance Resolved

Attributes	Data Type
CaseID	Varchar
Customer Number	Varchar
Fault duration	Varchar
Resolution Country	Varchar
Complaint Type	Varchar

Sample Key Attributes of Delivery Completed Table

Attributes	Data Type
No of Orders	Int
Installation Charges	Varchar
Order type	Varchar
Contract Number	Varchar

Attributes used for Modeling Clusm

Customer Tier	Product Services	Delivery State
Tier1	Voice	Delay
Others	Data	On-Time
	Managed Services and Others	Advanced

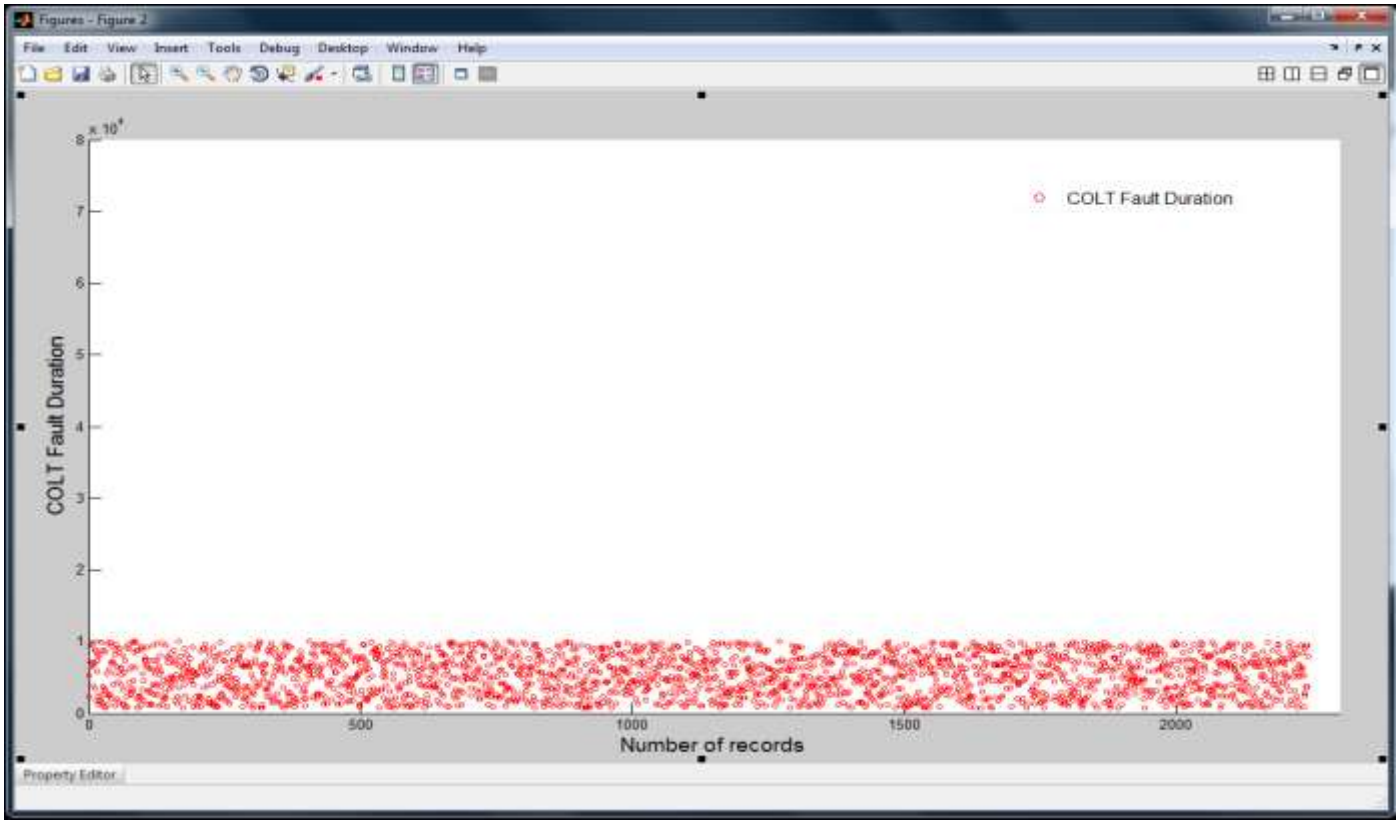
Data Preprocessing

- Null Value elimination
  - ✓ Is Empty() fn present in matlab was used to identify null values and row associated with it was removed
- Missing Value Elimination
  - ✓ Is NaN() fn present in matlab was used to identify missing values and row associated with it was removed
- Negative Value Elimination
  - ✓ Negative values cannot be part of the analysis being performed hence these values were identified and eliminated

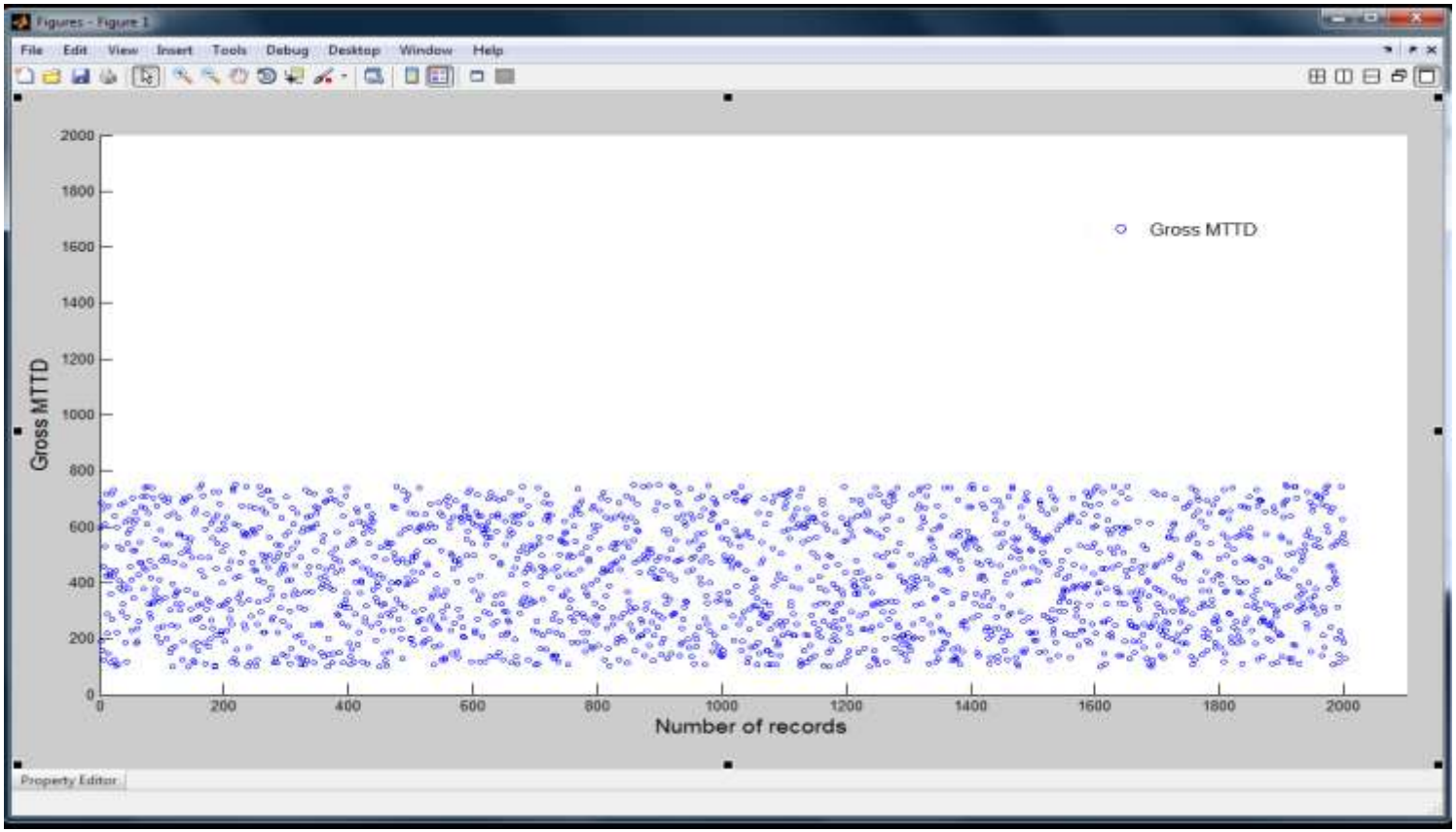
Methodology

- Sanitized data stored in MYSQL database
- Matlab R2012b was used for the analysis
- Query Builder of Matlab package was used to query data
  - ✓ Specific join queries on delivery completed and billing table based on customer number and timestamp were used
- Neural networks, Naïve Bayesian, Decision Trees have been reported to be used for analyzing telecom churn
- Current study used Stats tool box - Multivariate logistic Regression on the data
- The probabilities of churn and key drivers of churn for the two different customer namely tier 1 and non tier1 were found

Sample Screen Shots -1

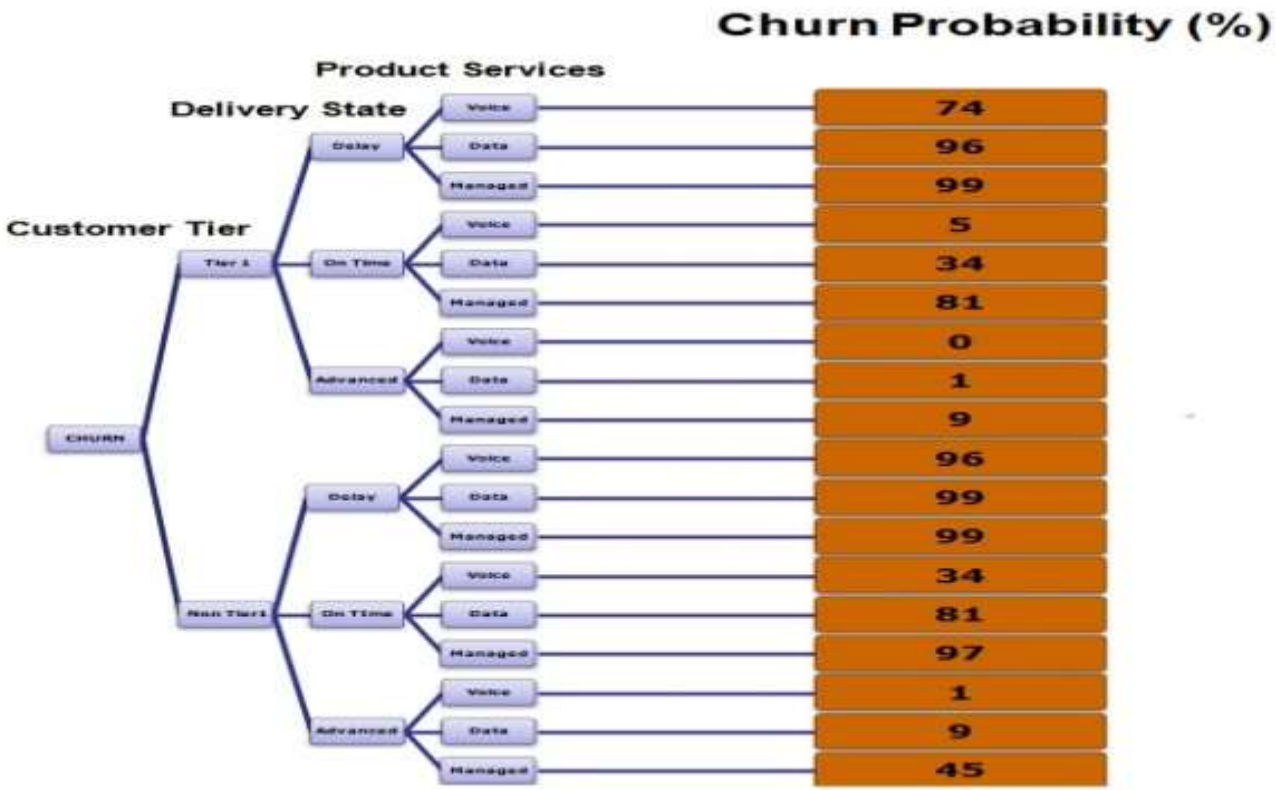


Sample Screen Shots - 2



Cognizant

Results



Churn = 1.054+ 2.1404 Customer Tier -3.8226*Delivery Status + 2.1404 * Product Services

Cognizant

Conclusion

- Delayed response whether in tier 1 or non tier 1 is the chief initiator of churn
- Quality of service could also a play major role even if the response /delivery state is on time
- Extraction of data using queries were time consuming
- Need to parallelize in order to make prediction faster
- Use of matlab production server is being explored

COGNIZANT

Thank you

Cognizant