Telecom Churn Prediction: a Big-Data Approach

António Fonseca Coelho – 54575 - MAEBD

afo.coelho@campus.fct.unl.pt

September 2020

Adviser:
Marta Lopes, FCT - NOVA

Co-Advisers:
João Pinto, Deloitte
João Lourenço FCT - NOVA

Context - What is churn?

- Churn is when a client quits

- From buying a product

- From using a service

Context - Why is it important to predict churn?

- In Telecom Industry aquiring a new client cost 5 times more than retaining an existing client

Aquiring a new client

Retaining an existing client

Problem – Predict churn

Problem - What is observable?

Static features

Time dependente features

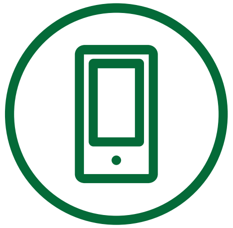
Problem - Challenges

- Static and time dependent features;
- Data comes from multiple sources;

Problem - Challenges

- Static and time dependent features;
- Data comes from multiple sources;

- Data is stored at different speeds;



Calling someone is something that might happen every day

Paying the bill is something that happens once a month

Problem - Challenges

- Static and time dependent features;
- Data comes from multiple sources;
- Data is stored at different speeds;
- Data volume exceeds 70 terabytes.

Problem - Challenges

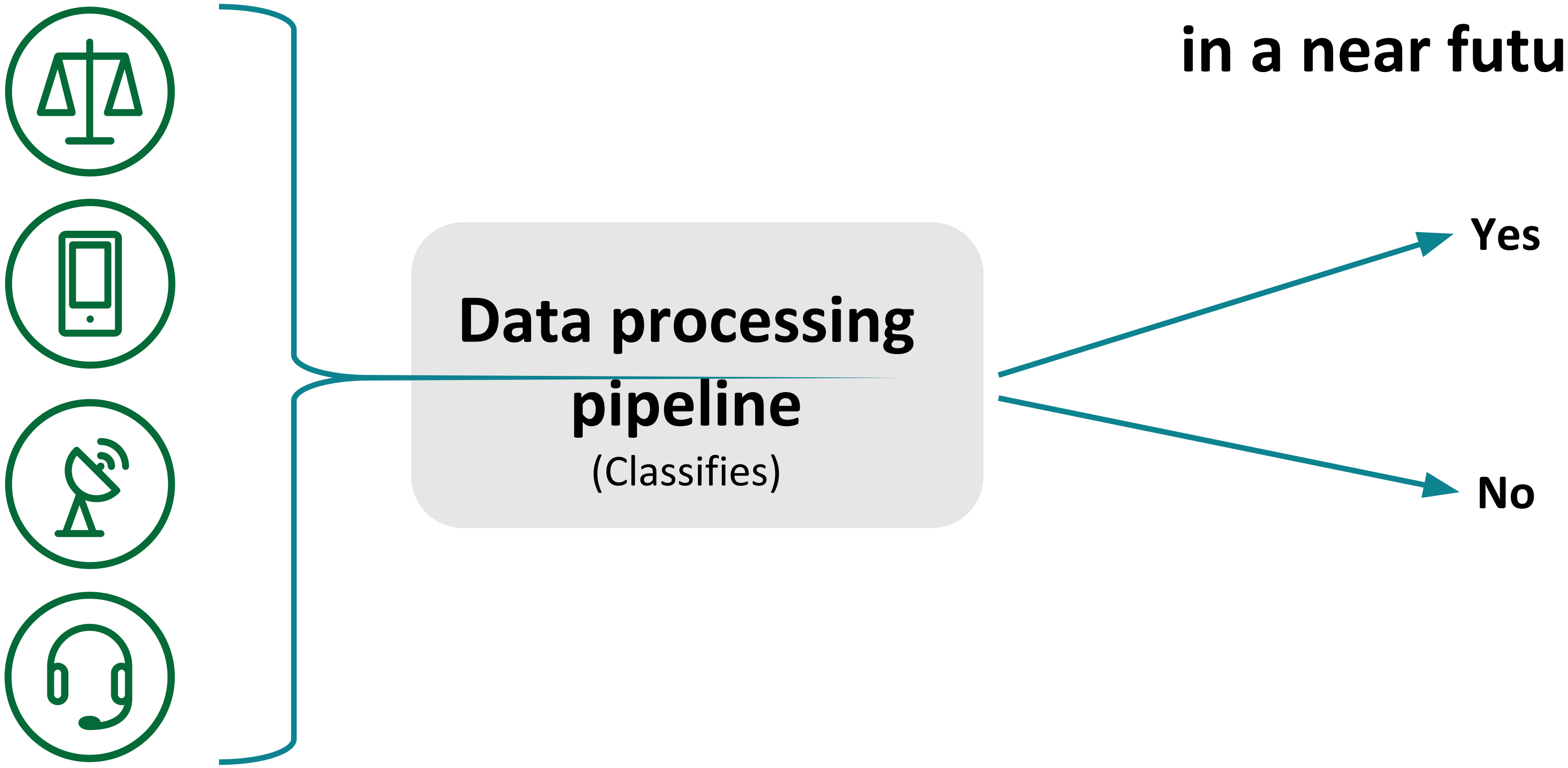
Multicollinearity

Sparsity

Imbalacing

Proposed approach

Will the client churn in a near future?



Index

- Context
- Problem
- Proposed Approach
- **Index**
- Related Work
- Proposed Approach
- Planning

Proposed approach steps

- Select a sample
- Create a framework which:
  - Data cleaning
  - Feature engineering
  - Feature selection
  - Finds solutions to imbalance problem
  - Models Churn
- Implement framework in the whole dataset

Related Work – Data Cleaning

1 Remove Unique Features (Contract ID…)

“Deleted features with identical values or missing values” (Sparsity)

“(…)deleted duplicated features(…)”(Multicollinearity)

“(…)features that have few numeric values(…)”(Sparsity)

A. Ahmad, A. Jafar, and K. Aljoumaa. “Customer churn prediction in telecomusing machine learning in big data platform.” In:Journal of Big Data6 (2019).doi:10.1186/s40537-019-0191-6.

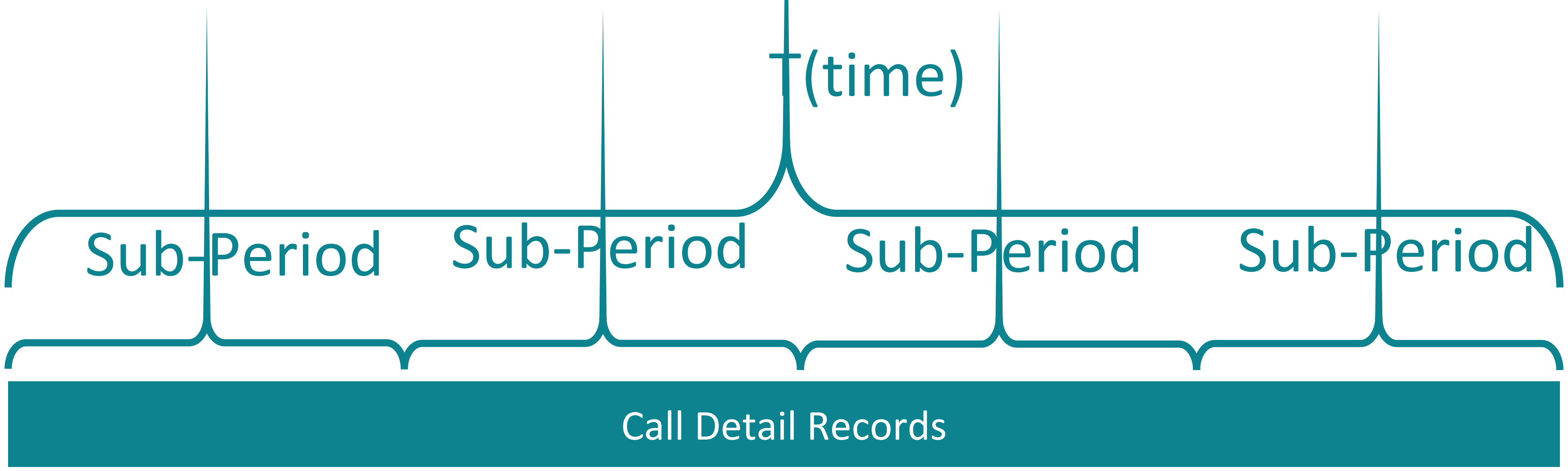
Related Work – Feature Engineering

- Static features – Most common value, Average…

- Dynamic features – Detect a change of behavior.

Related Work – Feature Engineering

- Static features – Most common value, Average…
- Dynamic features – Detect a change of behavior.

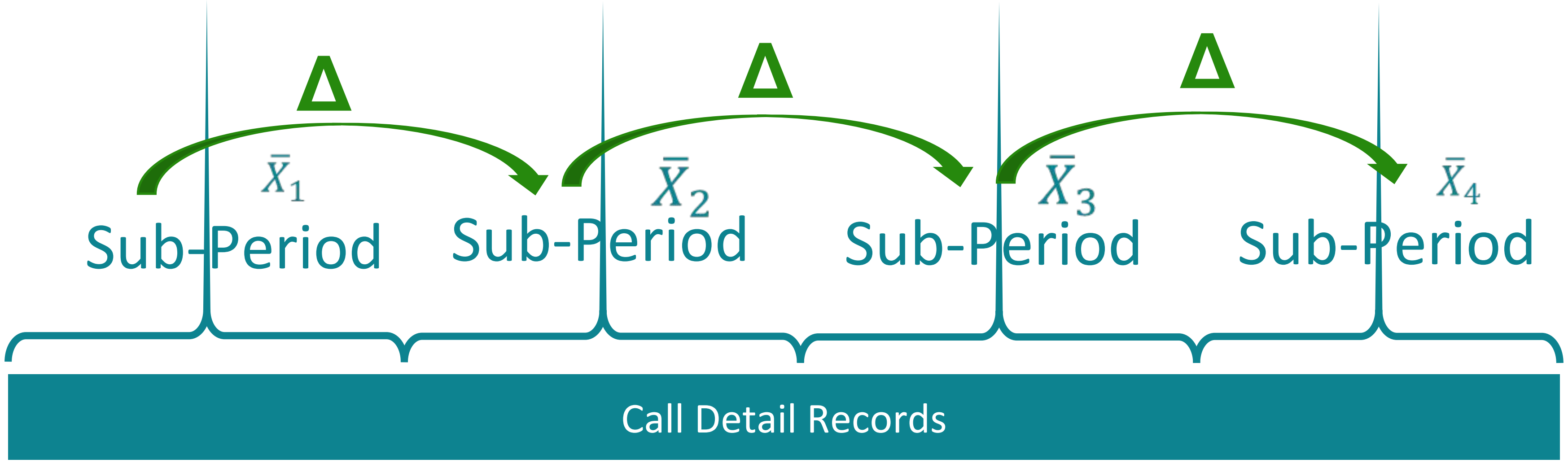


C. P. Wei and I. T. Chiu. **“Turning telecommunications call details to churn pre-diction: A data mining approach.”**
In: Expert Systems with Applications 23.2 (2002),pp. 103–112.

Related Work – Feature Engineering

- Static features – Most common value, Average…

- Dynamic features – Detect a change of behavior.



C. P. Wei and I. T. Chiu. “Turning telecommunications call details to churn pre-diction: A data mining approach.”
In: Expert Systems with Applications 23.2 (2002),pp. 103–112.

Related Work – Feature Selection

- Feature selection Methods

Filter methods

Wrapper methods

P. I. f. t. S. o. L. Langley and Expertise). "Selection of Relevant Features in MachineLearning."
In:In Proceedings of the AAAI Fall Symposium on Relevance. Vol. 184.1994, pp. 140–144.

Related Work – Imbalacing

- Typical balancing:
  - Over Sampling
  - Under Sampling
  - Class weight on cost function
  - Ensemble methods

Related Work – Imbalacing

- Typical balancing:
  - ~~Over Sampling~~
  - Under Sampling
  - Class weight on cost function
  - Ensemble methods

Related Work – Models used to predict churn

- Logistic regression
  - With Lasso regularization for embedded **feature selection**

- Decision trees models
  - Explainability of simple decision trees;
  - Random Forest is suitable for **high dimensional** data as it is an **ensemble** of trees trained in **subsets of features**

- Neural networks
  - Accepts **time-dependent** features directly as an input (e.g. LSTM, CNN.)
  - Reduce the amount of feature engineering effort

Index

- Context
- Problem
- Proposed Approach
- Index
- Related Work
- **Proposed Approach**
- Planning

Our data

- There is no single variable defining churn.

- But there is a variable that detects client activity.

Activity variable

Tru
e

False

Our data

- Activity variable - True or False

ID\mês	Jan	Feb	Mar	Apr	May	Jun
1	T	T	T	T	T	T
2	T	T	T	T	T	F
3	T	T	T	F	F	F
4	F	F	T	T	F	F
5	F	T	F	T	T	T

Proposed approach steps

- Select a sample
- Create a framework which:
  - Data cleaning
  - Feature engineering
  - Feature selection
  - Finds solutions to imbalance problem
  - Models Churn
- Implement framework in the whole dataset

Data cleaning

Proposed approach

- 1. Method to exclude repeated IDs and unique keys;
- 2. Method to identify one-time flags;
- 3. Method to treat personal information.

Feature Engineering

Proposed approach

- Static features:
  - Categorical Variables: Pick the last, most frequent value...
  - Numerical Variables: Average

- Time dependent features:
  - Custom groupby, e.g., calculate the difference between the last month with the average value

Feature Selection

Proposed approach

- Filter methods:
  - Univariate analysis (e.g., distribution)
  - Multivariate analysis (e.g., multicollinearity)

[16]A. J. Ferreira and M. A. Figueiredo. “Efficient feature selection filters for high-dimensional data.”
In:Pattern Recognition Letters33.13 (2012), pp. 1794–1804.issn: 01678655.doi:10.1016/j.patrec.2012.05.019.url:http://dx.doi.org/10.1016/j.patrec.2012.05.019.

Feature Selection

Proposed approach

- Filter methods:
  - Univariate analysis (e.g., distribution)
  - Multivariate analysis (e.g., multicollinearity)
- Wrapper Methods
  - Select a subset of variables
  - Based on the prediction model performance chose if this set is better than the other set
- Embedded Methods
  - Feature selection is done by the optimization algorithm of the prediction model (e.g. logistic regression with Lasso regularization).

Balancing

Proposed approach

- Typical balancing:
  - Under Sampling
  - Class weight on cost function
  - Ensemble methods
- Our case:
  - The distribution of the time as client for the churn class indicates that:
    - Most of clients that churn are recent clients.

Modeling

Proposed approach

- Logistic regression

  - Logistic Regression with lasso regularization – to understand if there is any feature that escaped the feature engineering process;

- Decision trees models

  - Decision Tree – to understand how the variables separate the data;
  - Random Forest – to test as a final model;

- Neural networks

  - Long Short Term Neural Network – to test as final model. Takes too much training with big-data;
  - Convolutional Neural Network – to test as final model. It is faster and outperforms LSTMs in some applications.

Which metrics?

Accuracy?

Which metrics?

~~Accuracy?~~

$$Accuracy = \frac{Nr\ of\ times\ model\ was\ correct}{Total\ nr\ of\ predictions}$$

If predictions are all = 0 => accuracy = 95%

Proposed Approach

Evaluation

- Confusion matrix

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

- Precision

How certain is a model when it classifies someone as churn?

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

- Recall

How many real churners is the model identifying?

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

- F1 Score

Harmonic average of precision and recall
A sense of how good the model is identifying churn in general

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Proposed Approach

Evaluation

- Client objective:
  - Maximize the investment.
- Client requirement:
  - Predictions must be precise.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

- **Precision**

How certain is a model when it classifies someone as churn?
Retain all customers

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

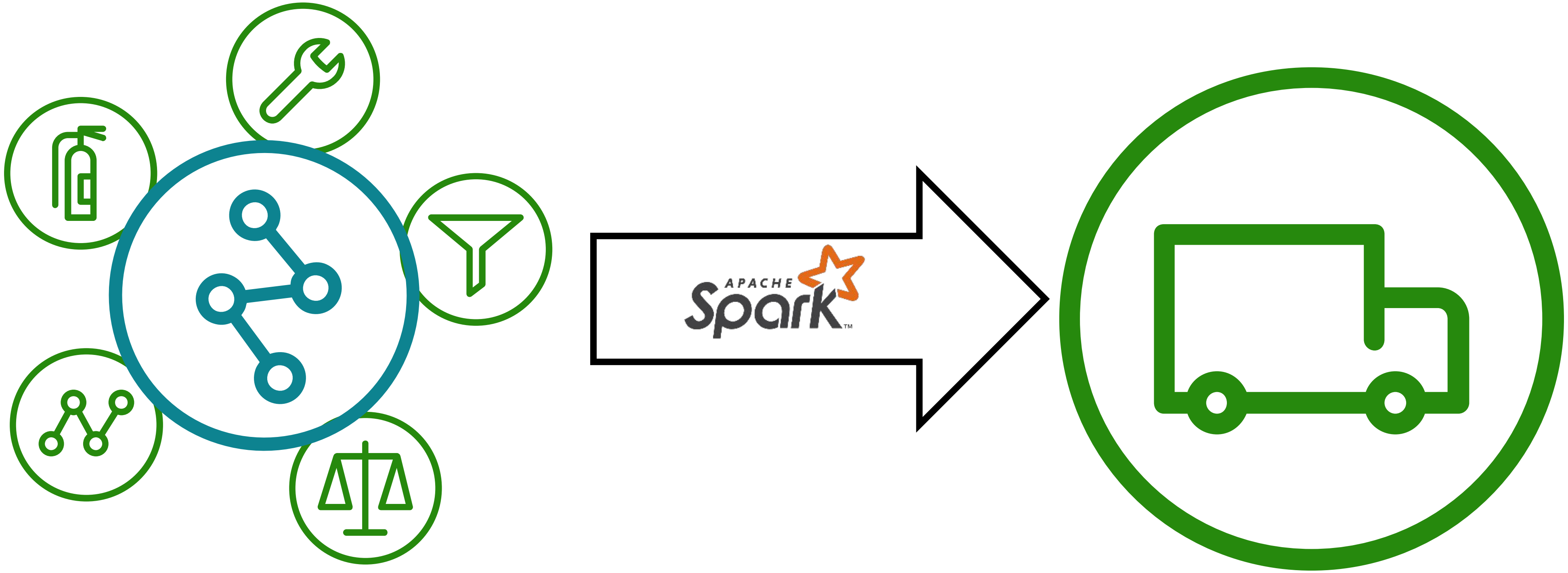
- **F1 Score**

Harmonic average of precision and recall
A sense of how good the model is identifying churn in general

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Proposed Approach

Implement framework in the whole dataset



Planning

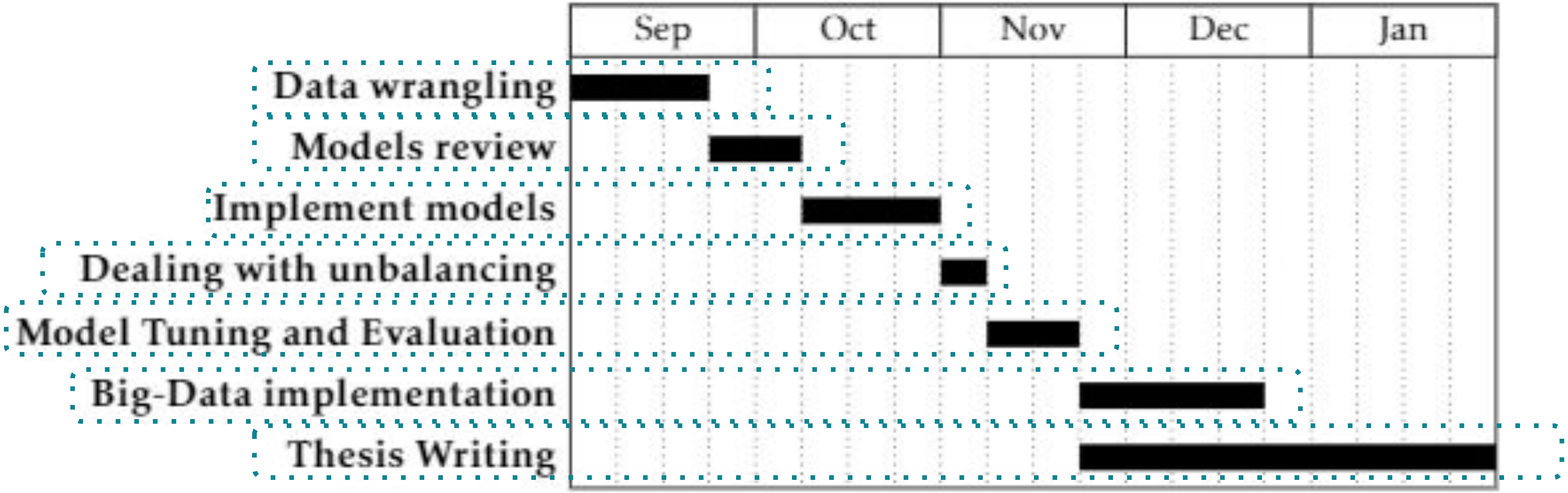
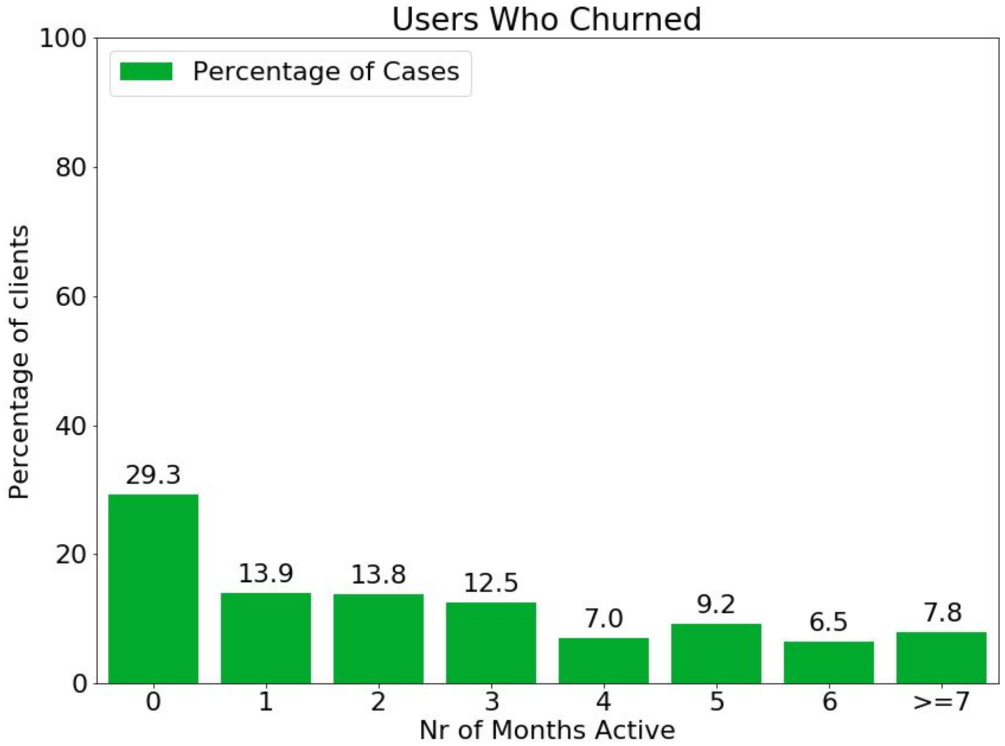
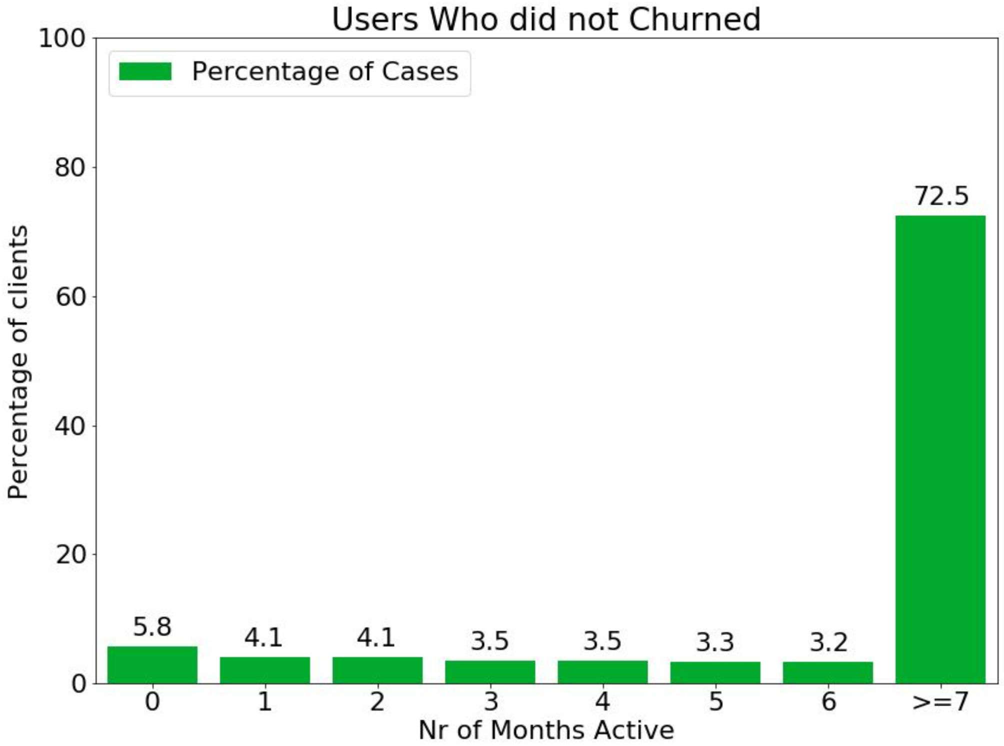


Figure 4.3: Work Plan Schedule.

Questions?


Users Who Churned
Percentage of Cases
Percentage of clients
Nr of Months Active
0
20
40
60
80
100
29.3
13.9
13.8
12.5
7.0
9.2
6.5
7.8
0
1
2
3
4
5
6
>=7


Users Who did not Churned
Percentage of Cases
Percentage of clients
Nr of Months Active
0
20
40
60
80
100
5.8
4.1
4.1
3.5
3.5
3.3
3.2
72.5
0
1
2
3
4
5
6
>=7