

# Natural Language Processing (NLP) in Business Applications

---

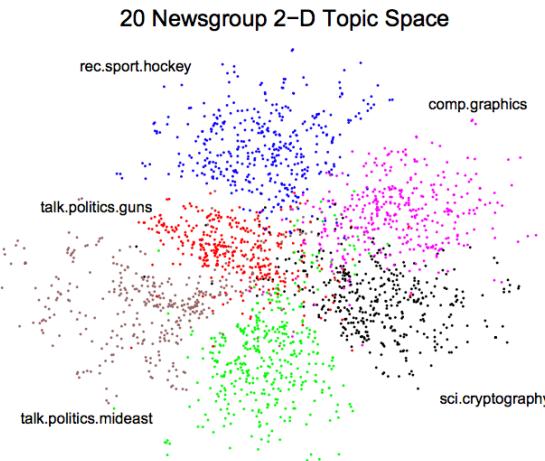
# I. Natural Language Processing (NLP)

- NLP is teaching computer to understand, process and analyze large amounts of human language data.
- In late 1980s, hand coded rules-based NLP evolved into automatic NLP with the emergence of statistical Machine Learning.
- Nowadays, NLP has been widely used in businesses:
  - Translation
  - Document Summary
  - Q&A
  - Text Classification
  - Sentiment/Emotion analysis

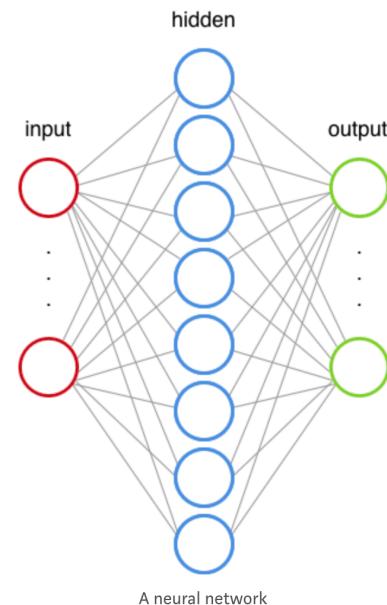
A repository to track the progress in NLP, including the datasets and the current state-of-the-art for the most common NLP tasks: <https://github.com/sebastianruder/NLP-progress>

# Text Classification

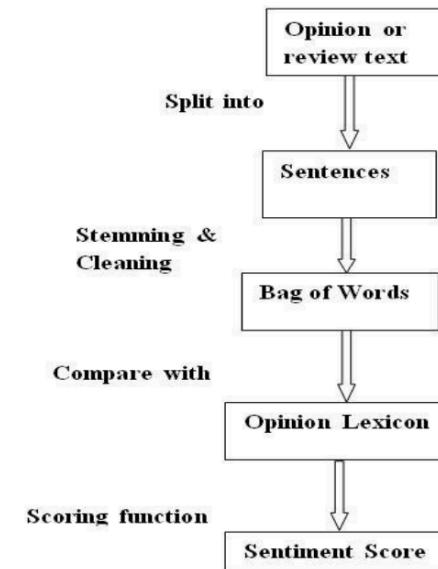
Unsupervised ML  
- Clustering



Supervised ML  
- Statistical Modeling



Lexicon based  
- Keyword Dictionary



*Semantic Hashing, Ruslan Salakhutdinov and Geoffrey Hinton (2007).*

*Big Picture Machine Learning: Classifying Text with Neural Networks and TensorFlow, Déborah Mesquita (2017).*

*Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study, Eman Younis (2015).*

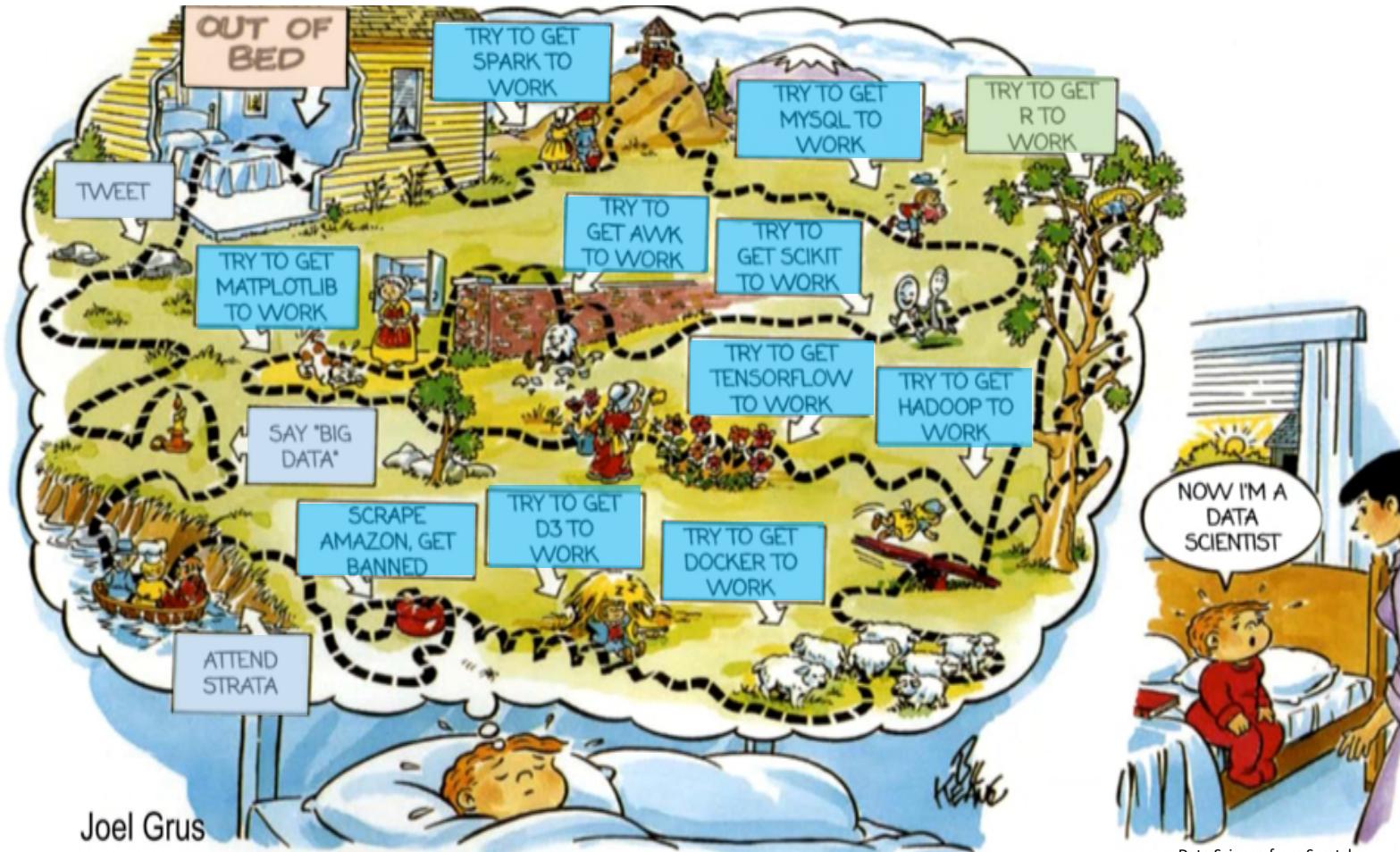
# Semi-Supervised NLP Classification

The information in the following slides are only shared for educational purpose for MSA8010 Spring.

Please advise your instructor before you forward the material to outside of the class.

Identify and quantify new product launch risk  
with Natural Language Processing

# 15 Ways to Be a Data Scientist



Work

Communication

Skill Sets

# Classification

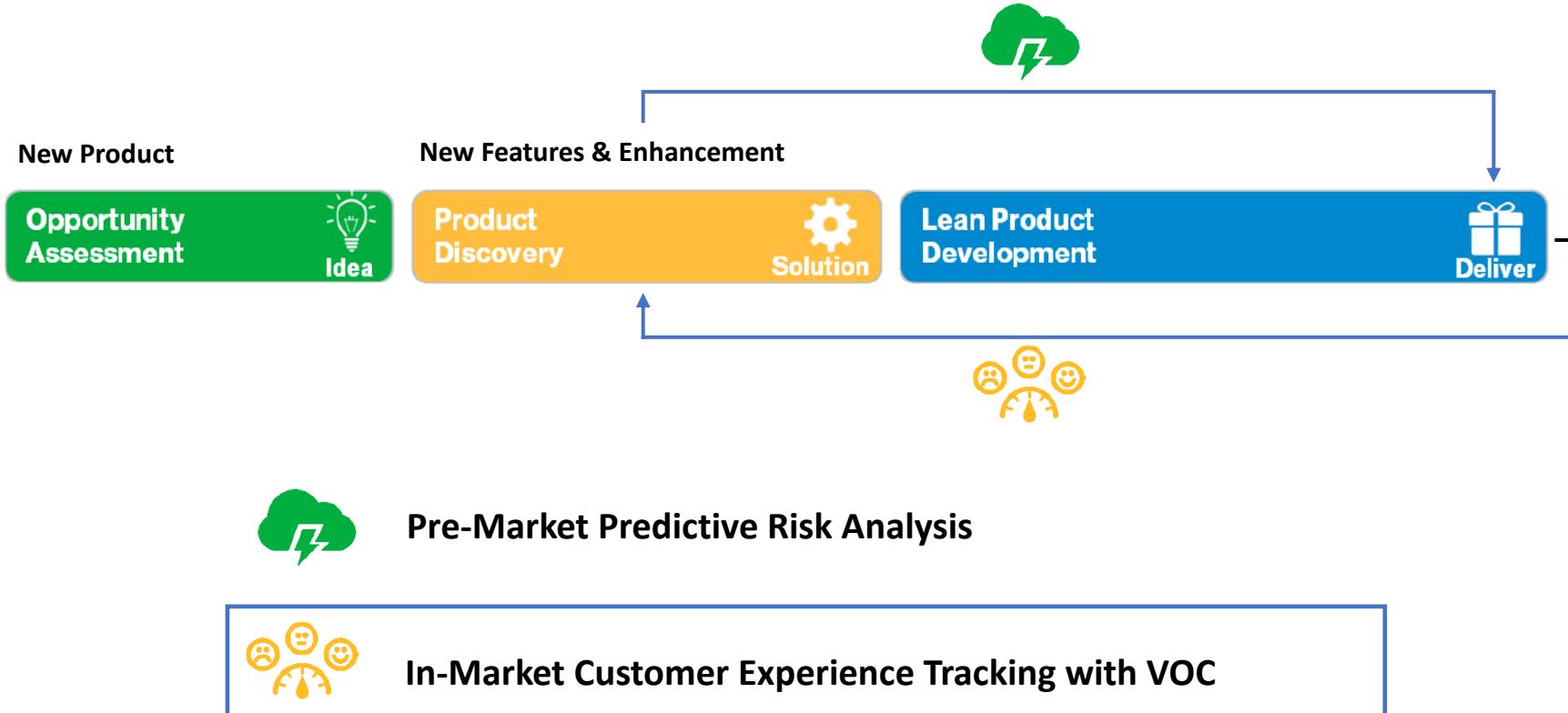


A lot of times, text analytics is a classification problem

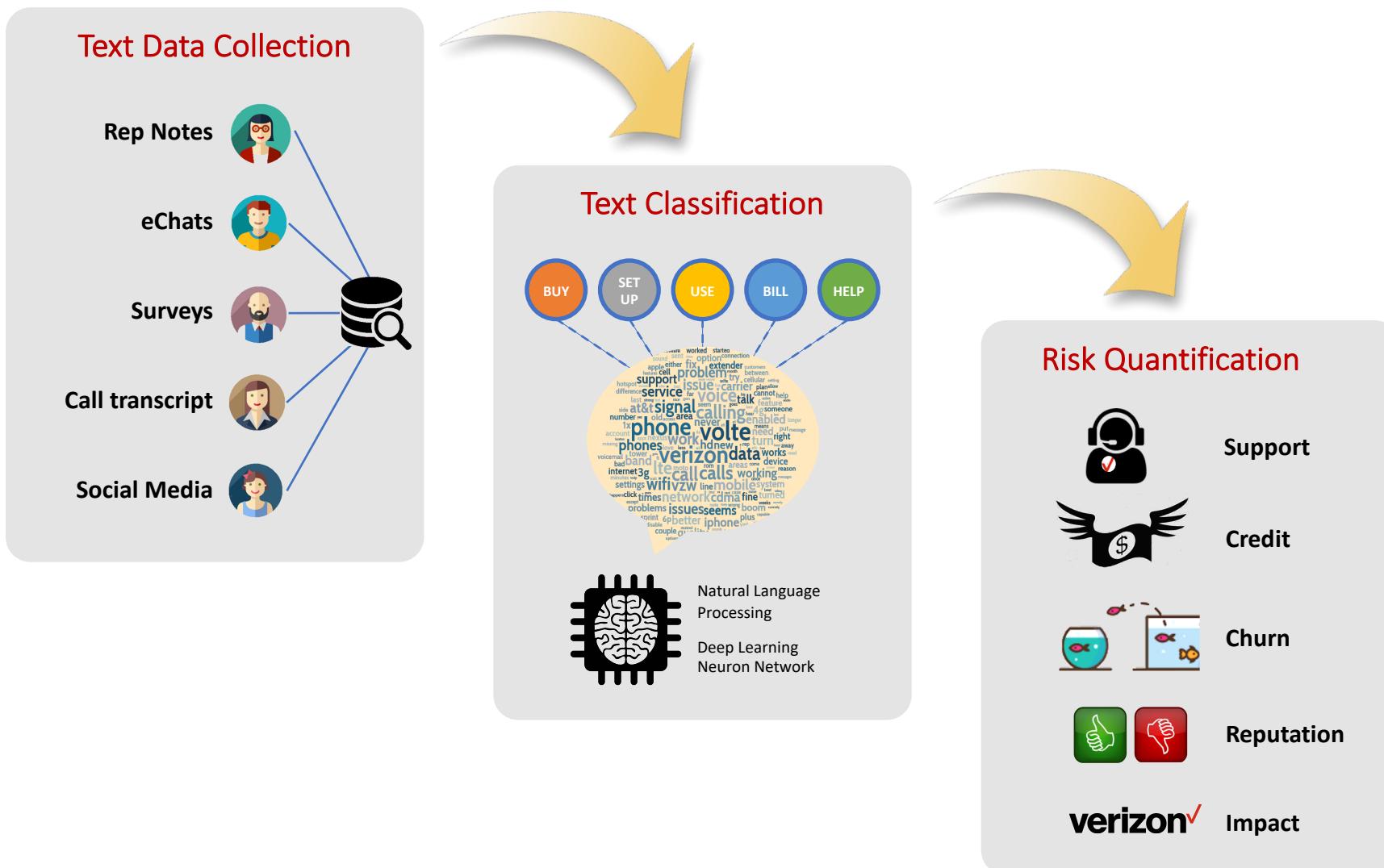
# Objective



# New Product Launch



# Solution Schema



# Text Classification

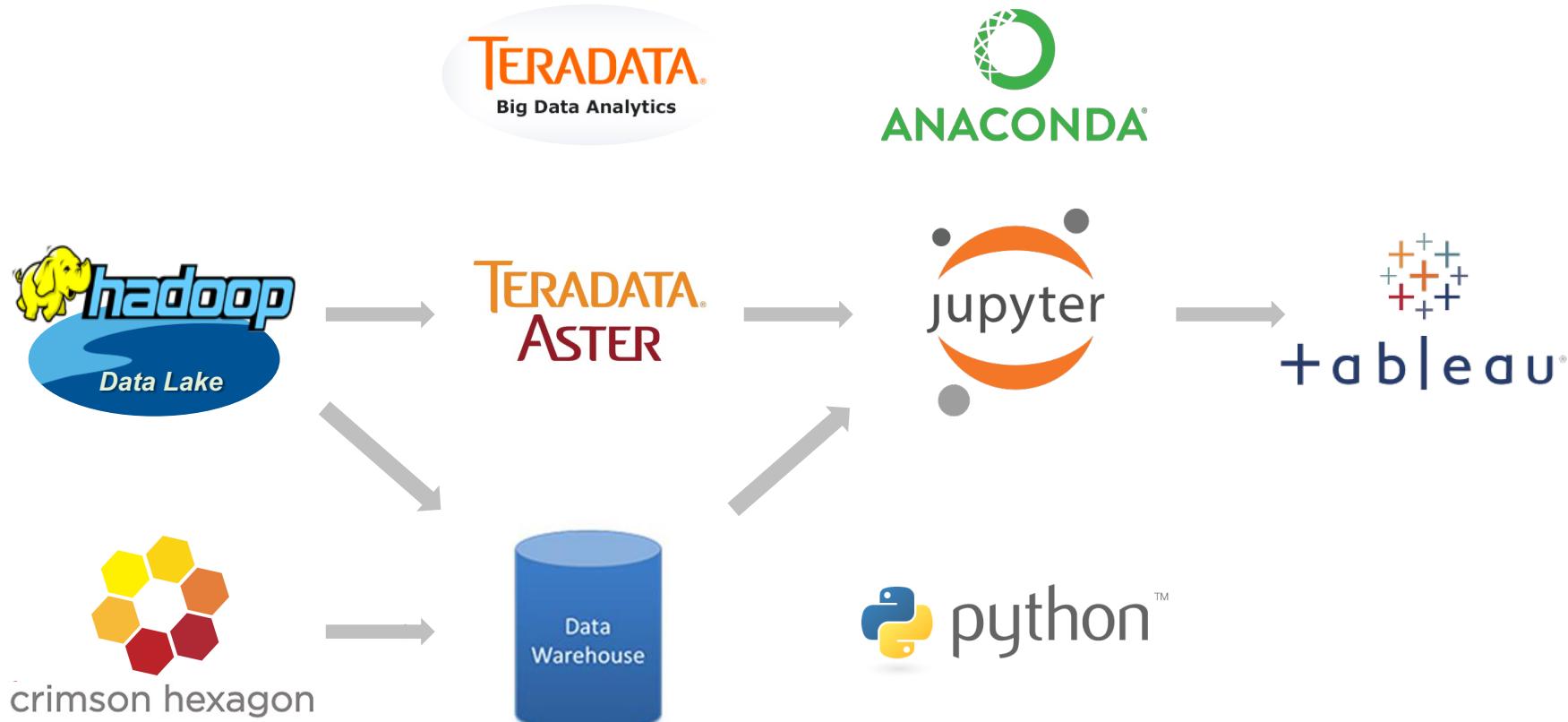
- **Unsupervised machine learning – clustering**
  - Pre-determined classification buckets
  - Hard to interpret clusters
- **Supervised machine learning – predictive modeling**
  - Labeling predictive targets is time consuming



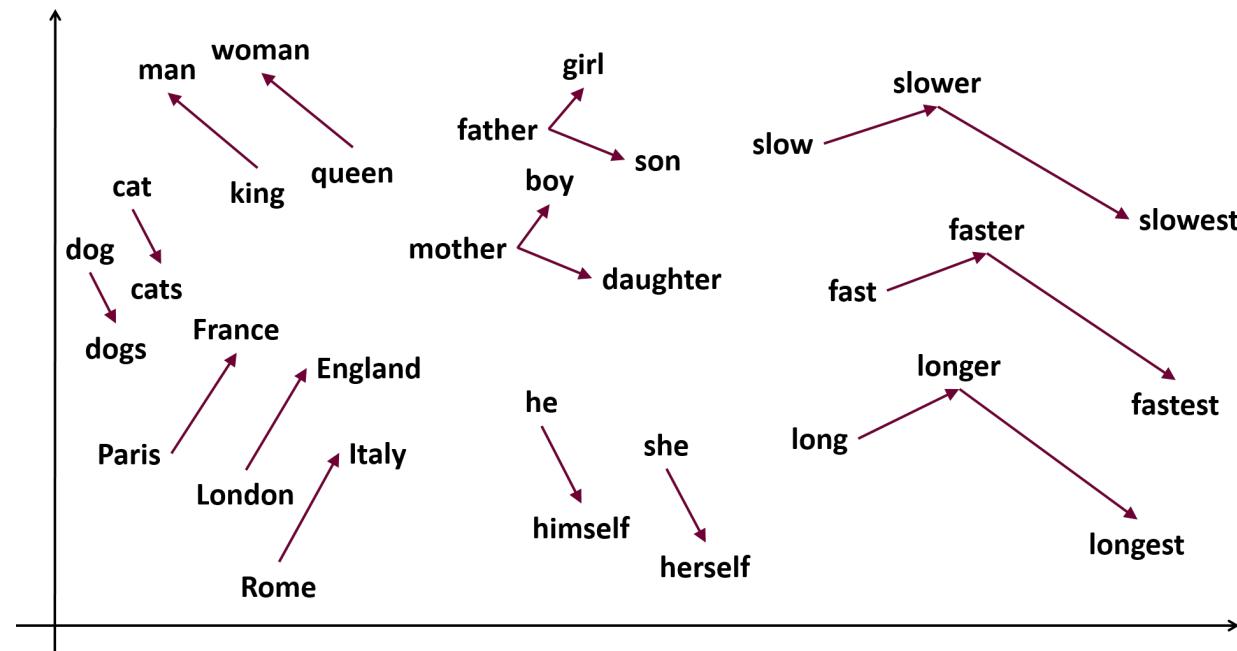
## Semi-supervised machine learning

- Key word dictionary for each customer journey phase
- Open source libraries such as word2vec, etc.
- Time efficient and easy to interpret

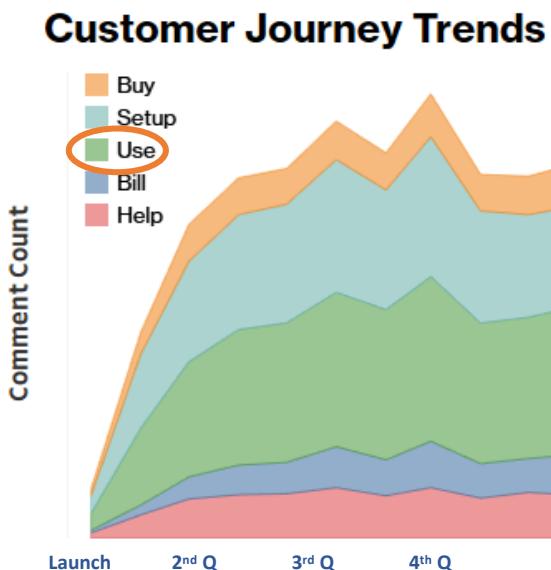
# Customer Journey Analytics Tools



# Word2Vec



# Results

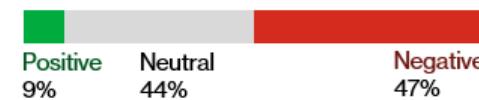


## Risk Quantification

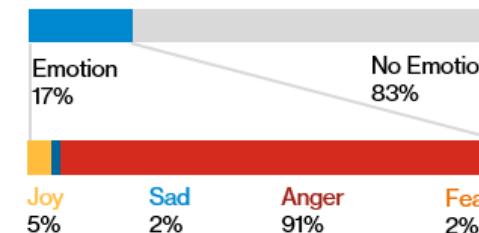
	Cust. Care	Credit Issued	Churn Loss	Total
Buy	\$618U	\$159U	\$196U	\$973U
Setup	\$2,167U	\$470U	\$328U	\$2,965U
Use	\$2,717U	\$528U	\$418U	\$3,663U
Bill	\$555U	\$170U	\$164U	\$168U
Help	\$829U	\$152U	\$155U	\$1,135U

U is a standardized unit.

## Sentiment



## Emotion

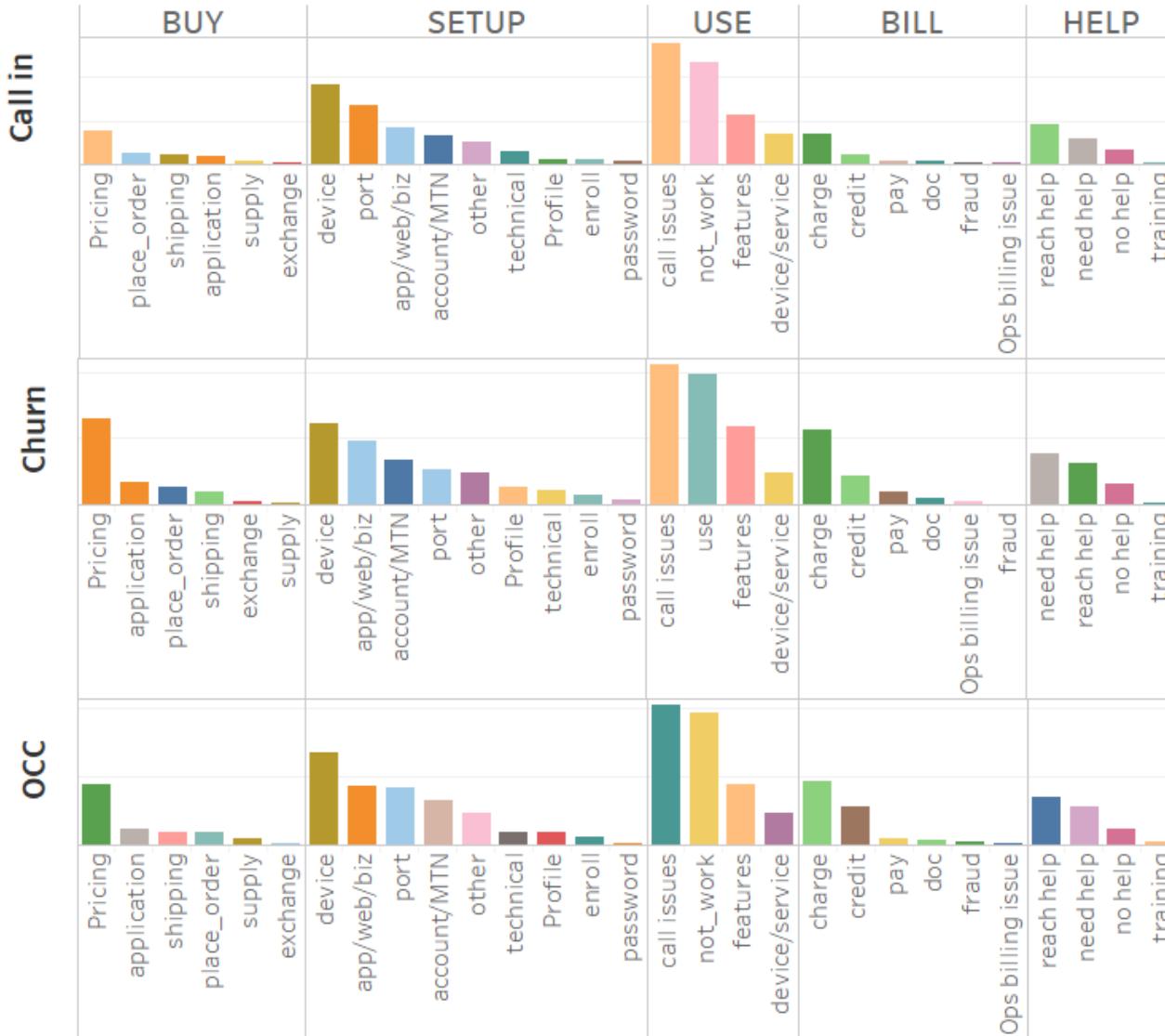


Yearly churn loss due to negative sentiment:  
\$2,549U

# Customer Pain Points for Use Phase

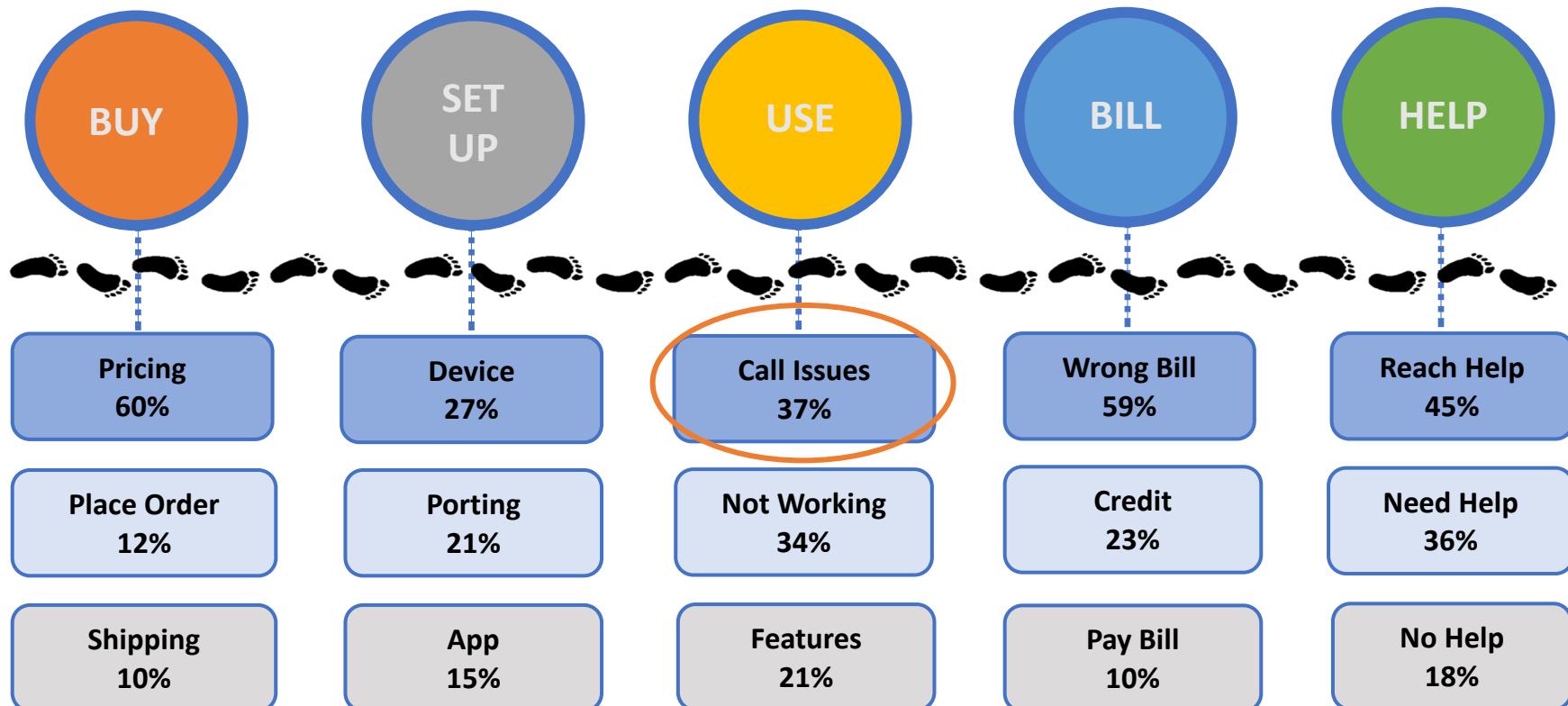


# Customer Pain Points



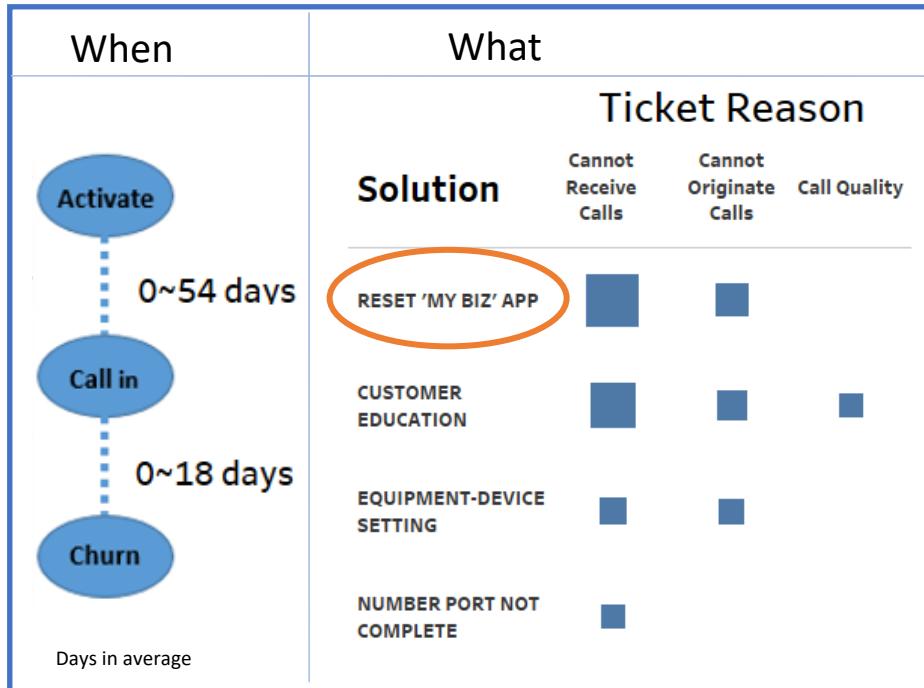
# Root Cause Analysis

## Call Reasons

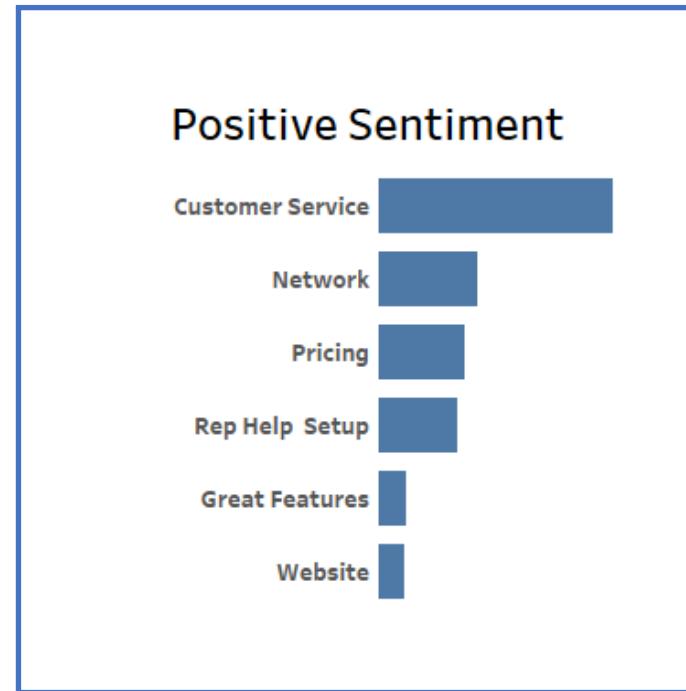


# Actionable Insights

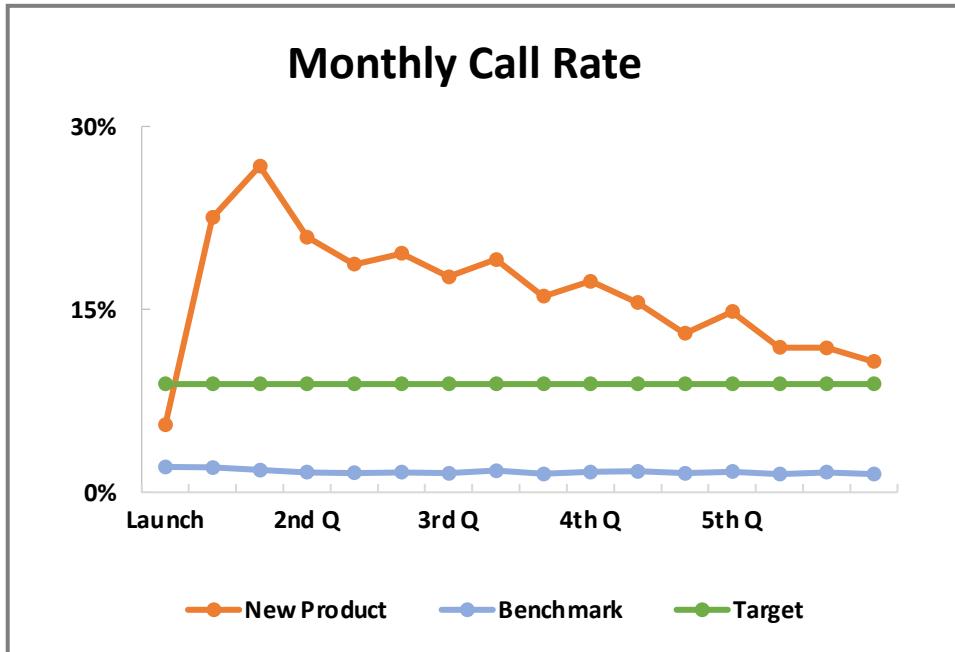
## Top Opportunity



## Top Strength



# Benefits



#### Target

- 40% average call rate reduction

#### Financial Benefit

- Cost saving on customer care operational cost was quantified.

# Python Demo

- Dataset:

<https://www.kaggle.com/CooperUnion/cardataset/data#>

- Code

<https://github.com/zhlli1/Genism-word2vec/blob/master/Genism%20Word2Vec%20Tutorial.ipynb>