Sentiment Analysis

BAX 453

Yuwen(Wendy) Ma
Qiuhong(Owen) Wei
Zhengjia(Vicky) Chen
Hanqiu(cici) Yu
Shaolong(Fred) Xue

AGENDA

1. Project Background
1. Business Problem
2. Business Understanding

2. Analytics Methodology
1. Data Collection and Preparation
2. Model Building
3. Insights & Interpretation

3. Business Outcome
1. Business Action
2. Deployment & Integration

Business Problem

1. Overall customer satisfaction level

3. Frequent complaints or dislikes mentioned by customers

2. Key features or aspects of the product that customers appreciate

4. Recommendations or endorsements made by customers in their reviews

Business Understanding

Marketing Strategy

Make effective positioning and resonates with potential customers

Product Development

Align with customer expectations

Customer Satisfaction

prioritize areas for improvement

Competitive Analysis
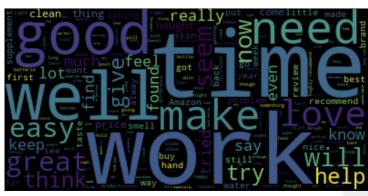
Highlight the unique selling points in advance

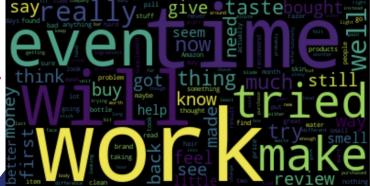# Data Collection and Preparation

reviewText	overall
This is a great little gadget to have around. ...	5
I would recommend this for a travel magnifier ...	4
What I liked was the quality of the lens and t...	4
Love the Great point light pocket magnifier! ...	4
This is very nice. You pull out on the magnifi...	5

>3

Positive Review
Label 1

Negative Review
Label 0

<3

Data Collection and Preparation

Bog of Words(BOW)
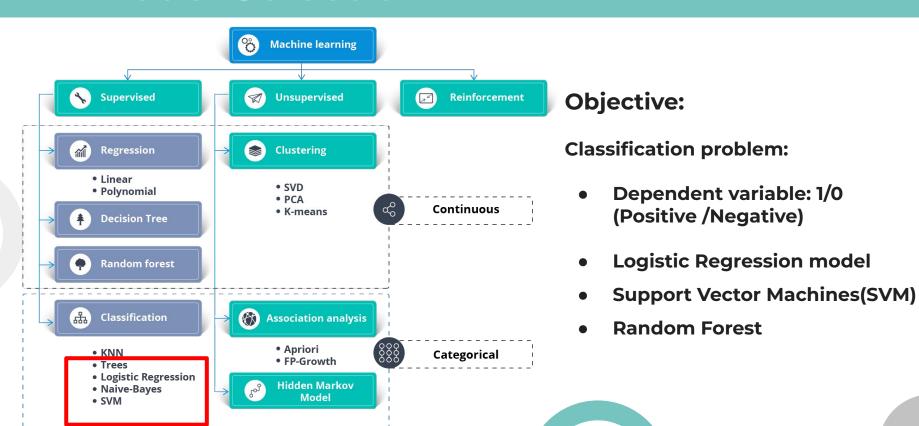
Vector indicate the frequency of words.

Term Frequency-Inverse Document Frequency (TF-IDF)

Reflect the importance of a term

N-Gram

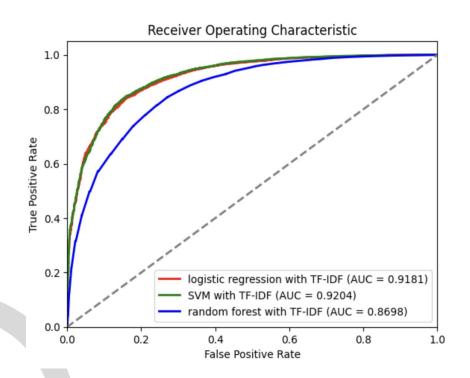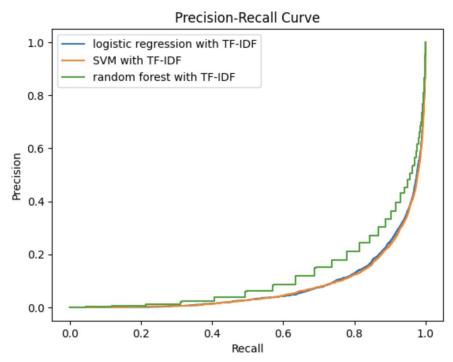Extract contiguous sequences of n items

Model Selection

Machine learning

Supervised · **Unsupervised** · **Reinforcement**

Regression
- Linear
- Polynomial

Decision Tree

Random forest

Clustering
- SVD
- PCA
- K-means

Continuous

Classification
- KNN
- Trees
- Logistic Regression
- Naive-Bayes
- SVM

Association analysis
- Apriori
- FP-Growth

Hidden Markov Model

Categorical

Objective:

Classification problem:

- **Dependent variable: 1/0 (Positive /Negative)**

- **Logistic Regression model**

- **Support Vector Machines(SVM)**

- **Random Forest**

Model Result

	Accuracy	Sensitivity	specificity	Precision for class 1	Precision for class 0
Logistic Regression (BOW)	0.92	0.97	0.48	0.91	0.67
Logistic Regression (TF-IDF)	0.92	0.99	0.33	0.91	0.82
Logistic Regression (N-gram)	0.91	0.99	0.2	0.89	0.76
SVM (BOW)	0.91	1	0.23	0.91	0.88
SVM (TD-IDF)	0.92	0.99	0.29	0.92	0.86
SVM (N-gram)	0.90	1	0.09	0.90	0.80
Random Forest (BOW)	0.89	1	0.07	0.90	0.88
Random Forest (TF-IDF)	0.90	1	0.07	0.90	0.91
Random Forest (N-gram)	0.89	0.98	0.20	0.91	0.59

Model Result

Both SVM with TF–IDF approach have the highest score based on the AUC-ROC curve.

Insights & Interpretation

NLP

Model

Insights

TF-IDF

- TF-IDF determines the relevance of the words.

SVM

- Suitable for high-dimensional feature
- Less prone to overfitting
- Inherently designed for binary classification

Usage

- High performance score for both positive and negative reviews.

Business Action

Product Innovation & Competitive Benchmarking

- Improve products based on reviews
- Compare features of competing items

Targeted Marketing Strategy

- Market popular features
- Amplify with sponsorships

Proactive CRM

- Damage control
- Retain customers

Deployment & Integration

Product Innovation

Utilize the sentiment analysis

If customers frequently praise a competitor's product for its eco-friendly packaging, the client could consider redesigning their package.

Marketing Strategy

Targeted Marketing by Positive Feedback

If customers appreciate fragrance-free or works well for sensitive skin, these features should be emphasized in the product descriptions and marketing materials.

Customer Relationship Management

Negative Reviews Analysis

Company can reach out to customers proactively. This could involve providing a refund, offering a replacement product, or providing assistance to resolve their issues.

THANKS

Do you have any questions?