



Literature Survey, Collected Data, Proposed Methodology

Team - 14:

20101120 Sadiul Arefin Rafi

20101291 Pritam Barua

20101386 Nafisa Muhammad

21301232 Irfanul Hoque



Literature Survey

- Elli, Maria, and Yi-Fan's study on business analytics using SVM and MNB.
- Different classifiers for precision and recall.
- Sentiment analysis using Amazon's Kindle book reviews.
- Classifier evaluations displayed via statistical charts without accuracy measurements.
- Bag of words model for predicting product ratings.
- Use of the naive Bayes classifier with simpler algorithms for easy comprehension.



Collected Data

- Dataset sourced from Kaggle.
- Includes Amazon product reviews with details like 'Reviewer ID', 'ASIN', 'Reviewer Name', 'Review title', and more.
- Focus on gadgets like Kindle, Bluetooth speaker, TV, and Tablet.



Proposed Methodology

- Convert ratings into binary classification: Positive (≥ 3) as 1 and Negative (< 3) as 0.
- Tokenize reviews, remove stop words, and apply POS tagging.
- Lemmatize using WordNet Lemmatizer and NLTK POS tagger.
- Train various classifiers on the processed dataset.



Thank You



Data Analysis, Prototype Implementation, Result Analysis

Team - 14:

20101120 Sadiul Arefin Rafi

20101291 Pritam Barua

20101386 Nafisa Muhammad

21301232 Irfanul Hoque



Data Analysis

- Breakdown of text into tokens.
- Removal of stop words for better accuracy.
- Lemmatization to determine the root of words.
- POS tagging using NLTK library.



Prototype Implementation

- Used supervised algorithms to understand word sense.
- Applied various classifiers like decision trees, logistic regression, naive Bayes, MNB, SVM, and random forests.
- SVM chosen due to its highest accuracy.



Result Analysis

- Logistic Regression: 93.70%
- Multinomial Naive Bayes: 93.45%
- SVM (pre-optimization): 93.94% and post-optimization: 94.08%
- Decision Tree: 90.18%
- Random Forest: 93.46%



Thank You