

# Sentiment Analysis for Movie Reviews

Yuqi Jiang 001638771  
Yuxiao Chen 001618697

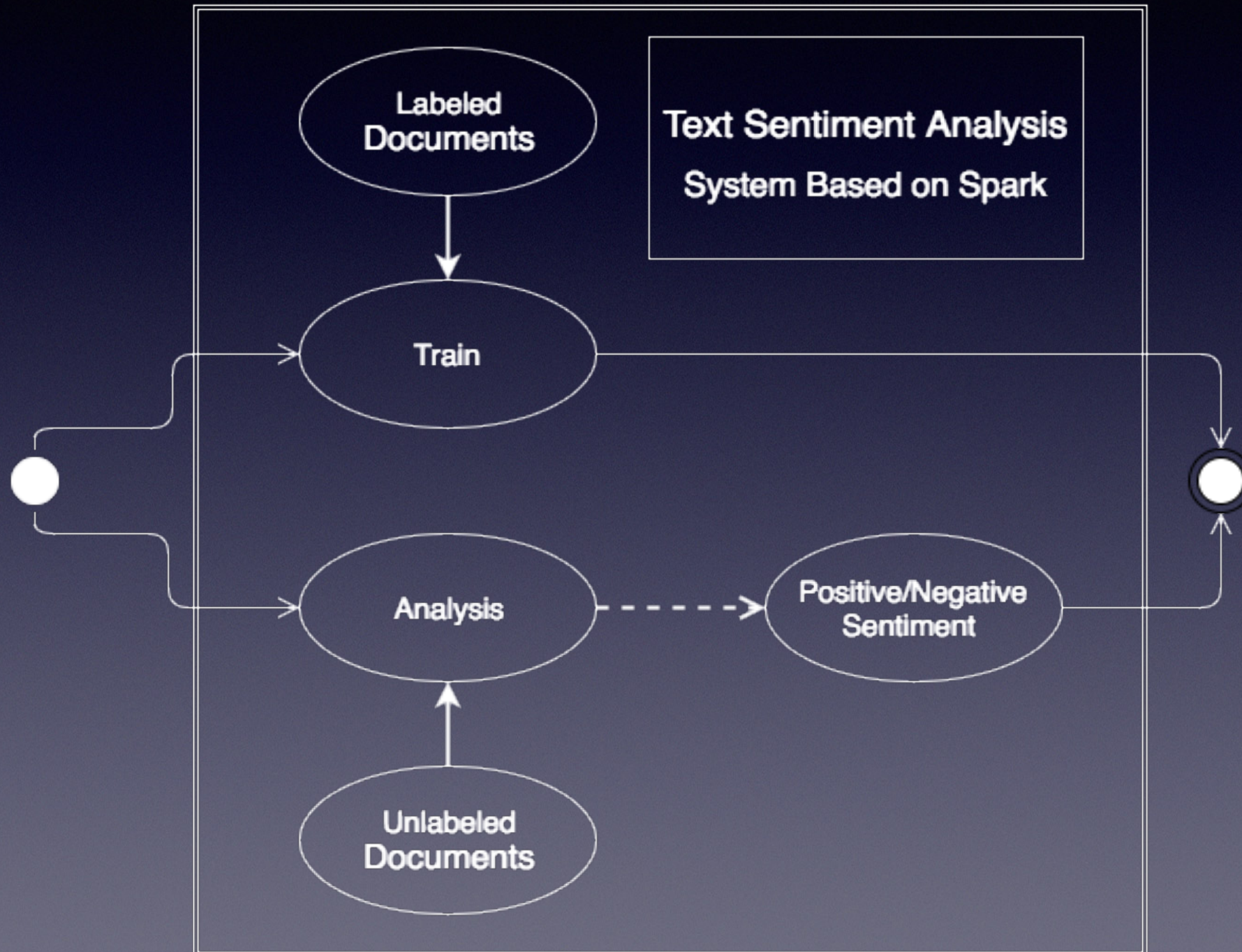


# Goal of the Project

- The main goal of this project is to predict the sentiment of a movie review using Scala and Spark



# Use Cases



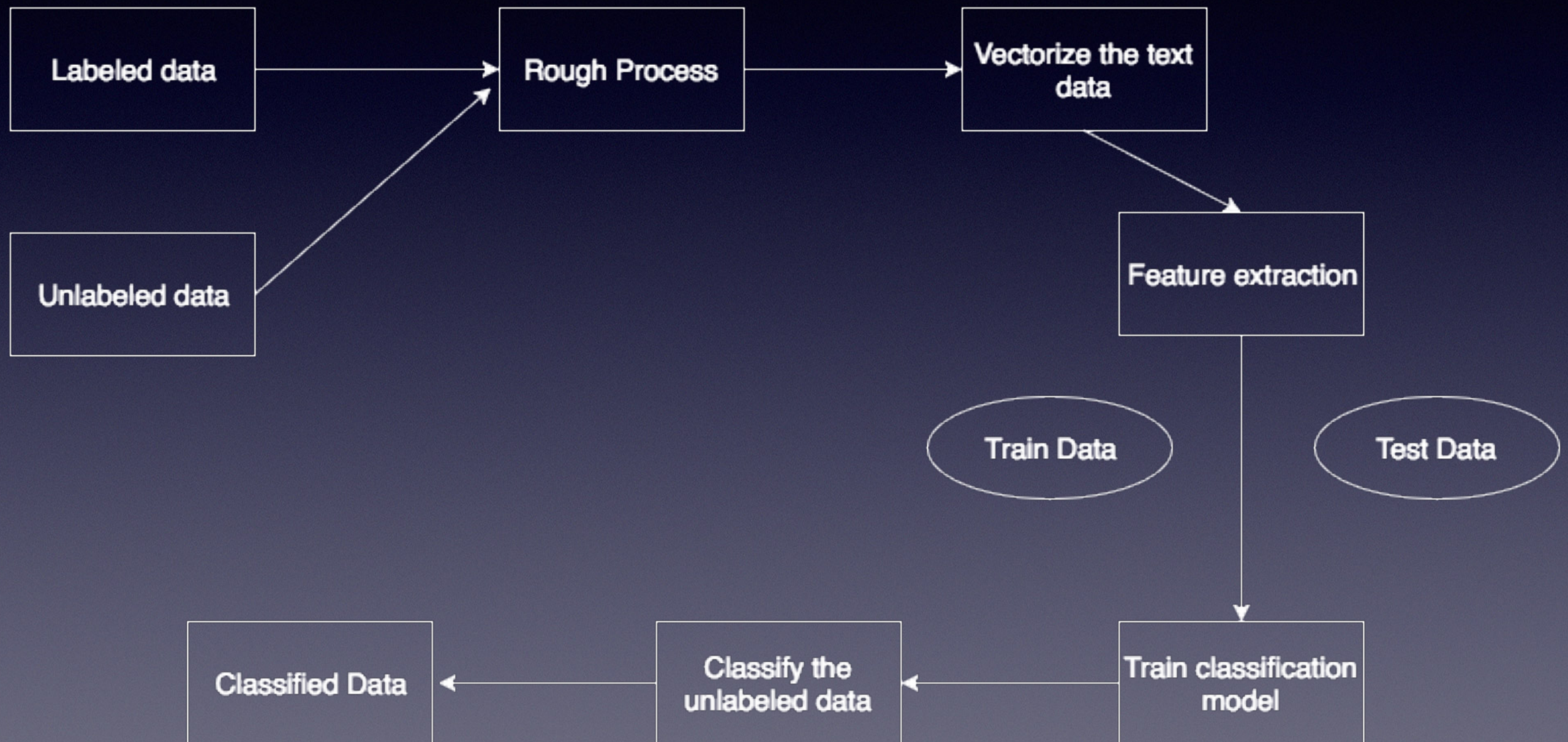


# Methodology

- Rough Process ——— RDD
- Feature Extraction and Vectorize Text Data———  
TF-IDF
- Train Classification Model ——— Naïve Bayes



# Methodology





# Data Sources

- <http://ai.stanford.edu/~amaas/data/sentiment/>
- 50K Labeled Data for Training and Testing
- Unlabeled Data for Analysis



# MileStones

- Rough Process — 11/7
- Feature Extraction and Vectorize Text Data — 11/15
- Train Classification Model — 11/20
- Test and Evaluation — 12/1



# Code Repository

- <https://github.com/Jaackey/CSYE-7200-Final-Project>



# Acceptance Criteria

- Give Prediction Based on Input
- Accuracy Achieves 60%