

NLP For Disaster Tweets Analysis

Group 14

Chuqin Wu

Cewen Zhu

Ning Yang

Rui Hu

Yaozhong Huang

Yilong Chen



Project Overview



Preprocessing

Tokenization,
Lemmatization,
Stopword Removal



Model Development

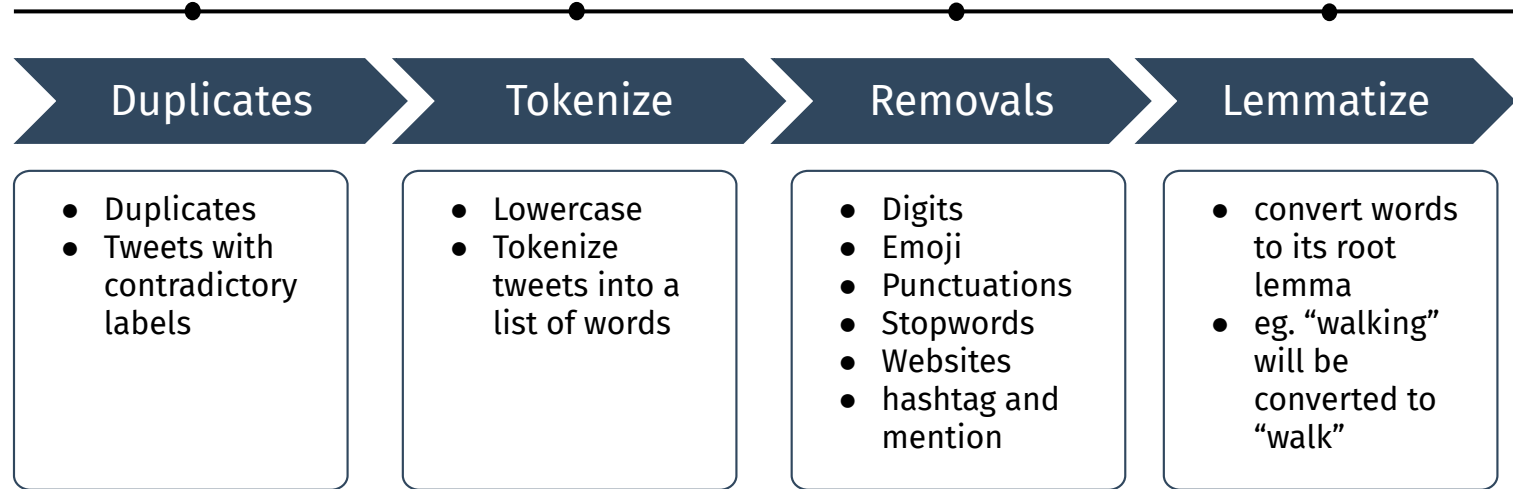
SVM and BERT



Optimization

Reduce loop, cython,
multiprocessing

Preprocessing



SVM

01

Word Vectorization

- Bag of Words
 - Python Counters

02

Model Development

- Pegasos SVM
 - Stochastic GD to improve efficiency

03

Hyperparameter Tuning

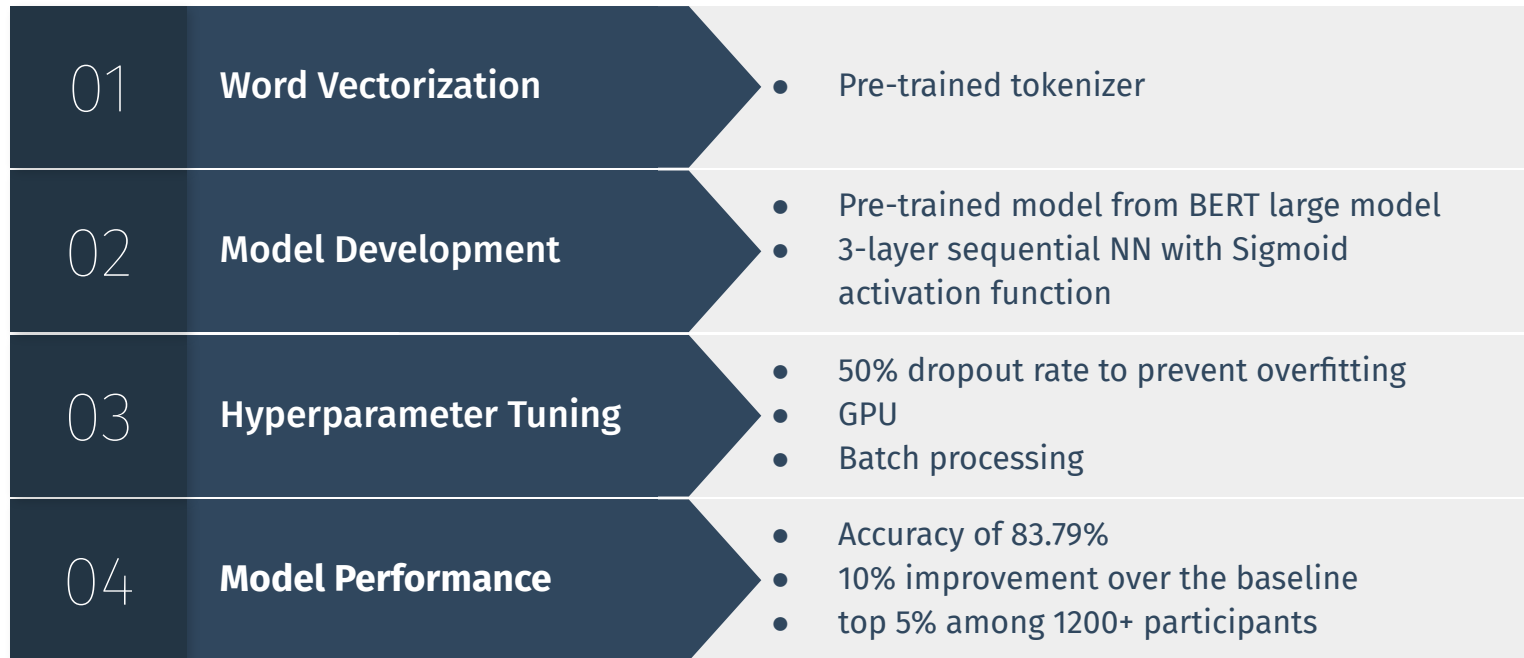
- Optimal $\lambda = 0.0078$

04

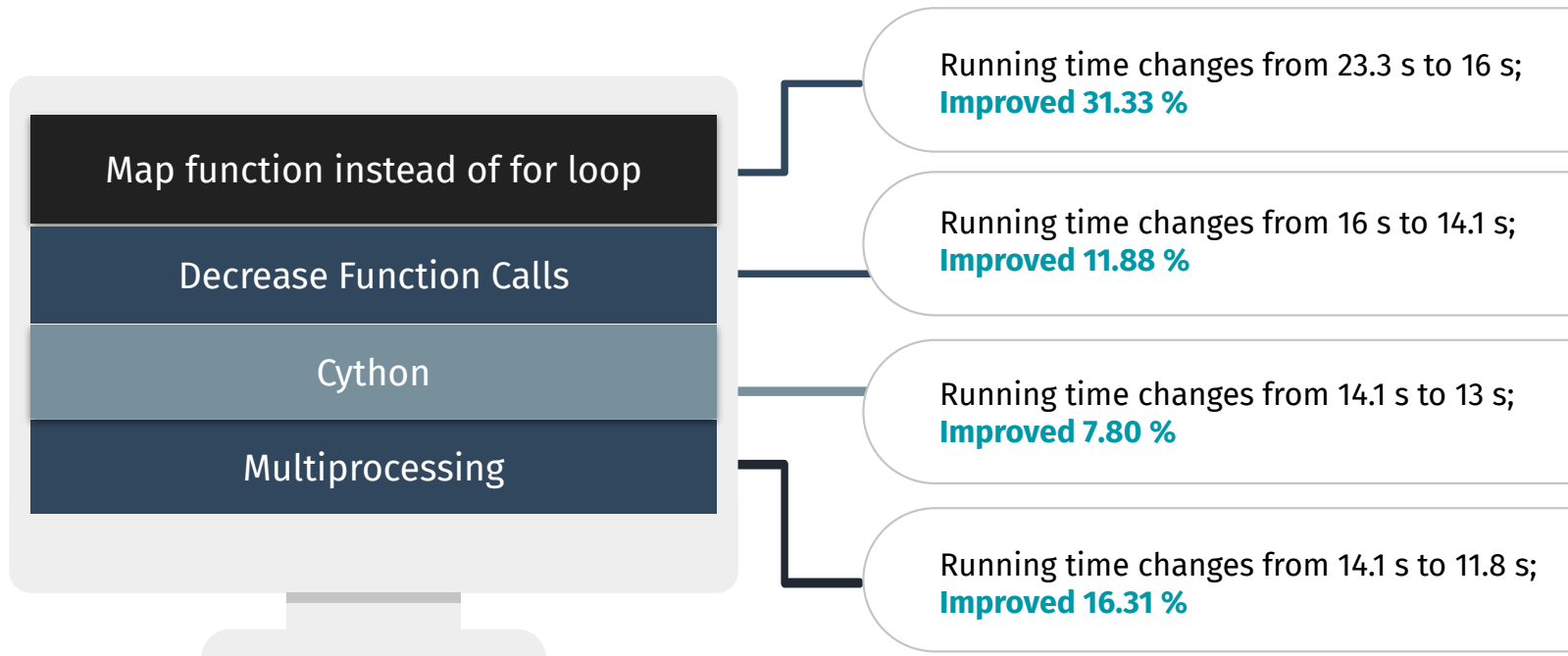
Model Performance

- Accuracy of 76.69%

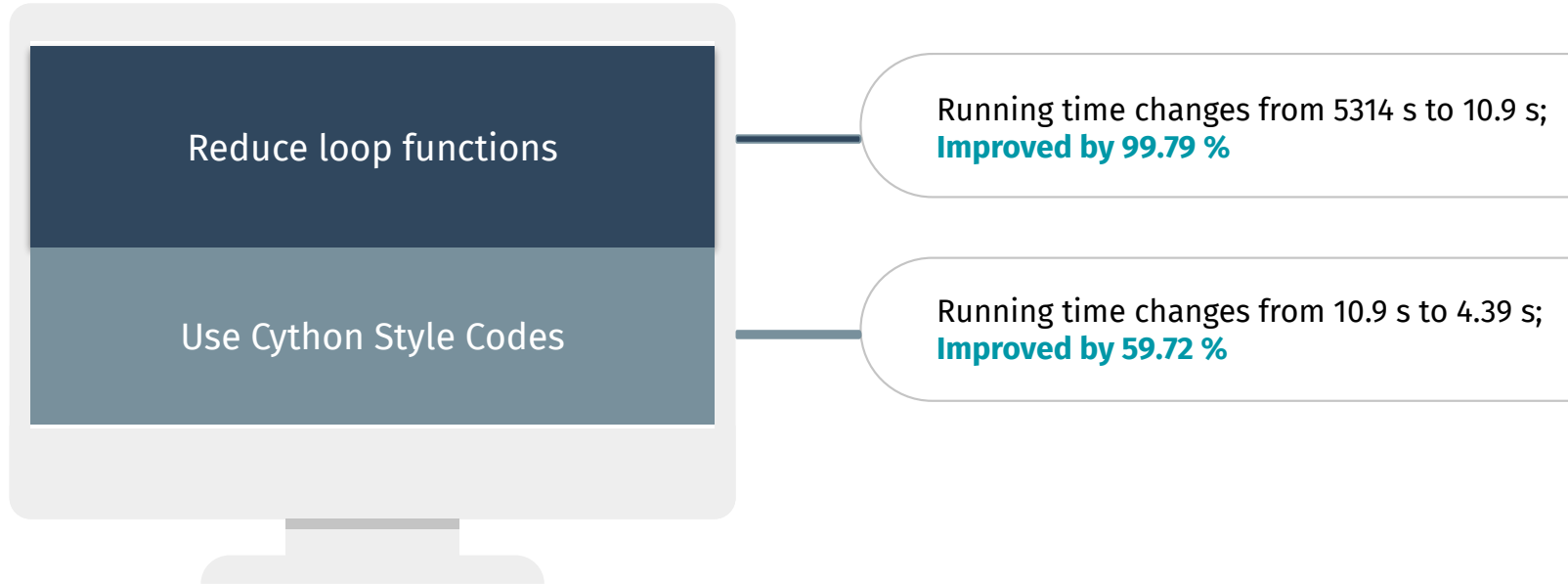
BERT



Optimization of Preprocessing



Optimization of SVM



Overall, SVM running time has improved 99.91%



Thank you!