

NLP For Disaster Tweets Analysis

NAME :- NIKHIL SAVALIYA

DIGICROME ACADEMY



Project Overview



Preprocessing

Tokenization,
Lemmatization,
Stopword Removal



Model Development

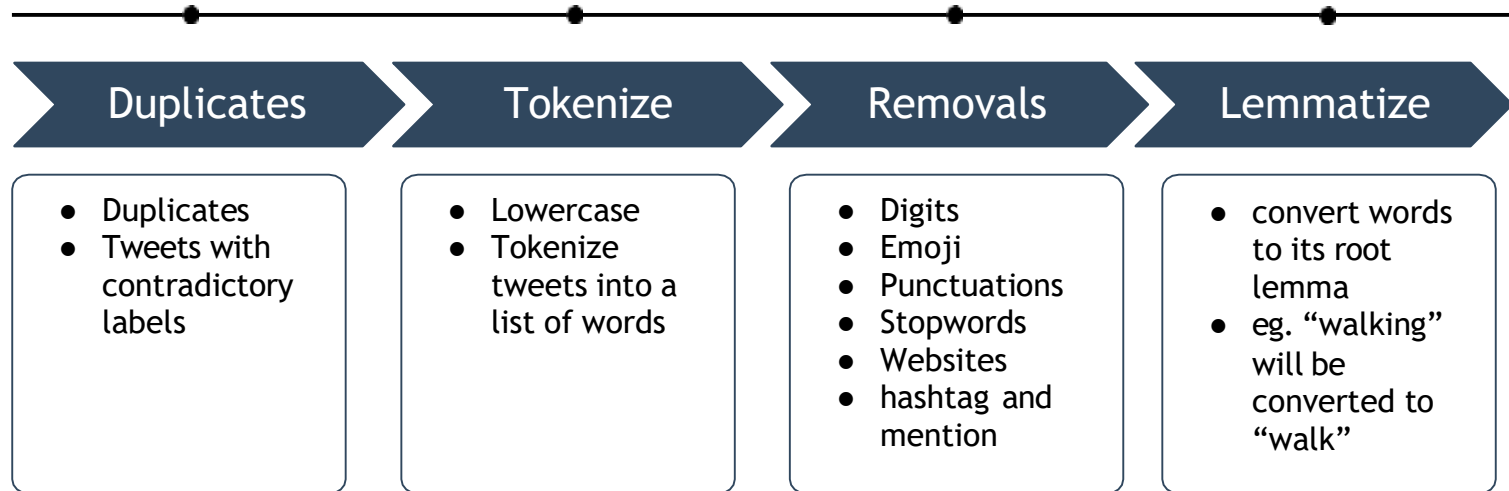
SVM and BERT



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

01

Word Vectorization

- Pre-trained tokenizer

02

Model Development

- Pre-trained model from BERT large model
- 3-layer sequential NN with Sigmoid activation function

03

Hyperparameter Tuning

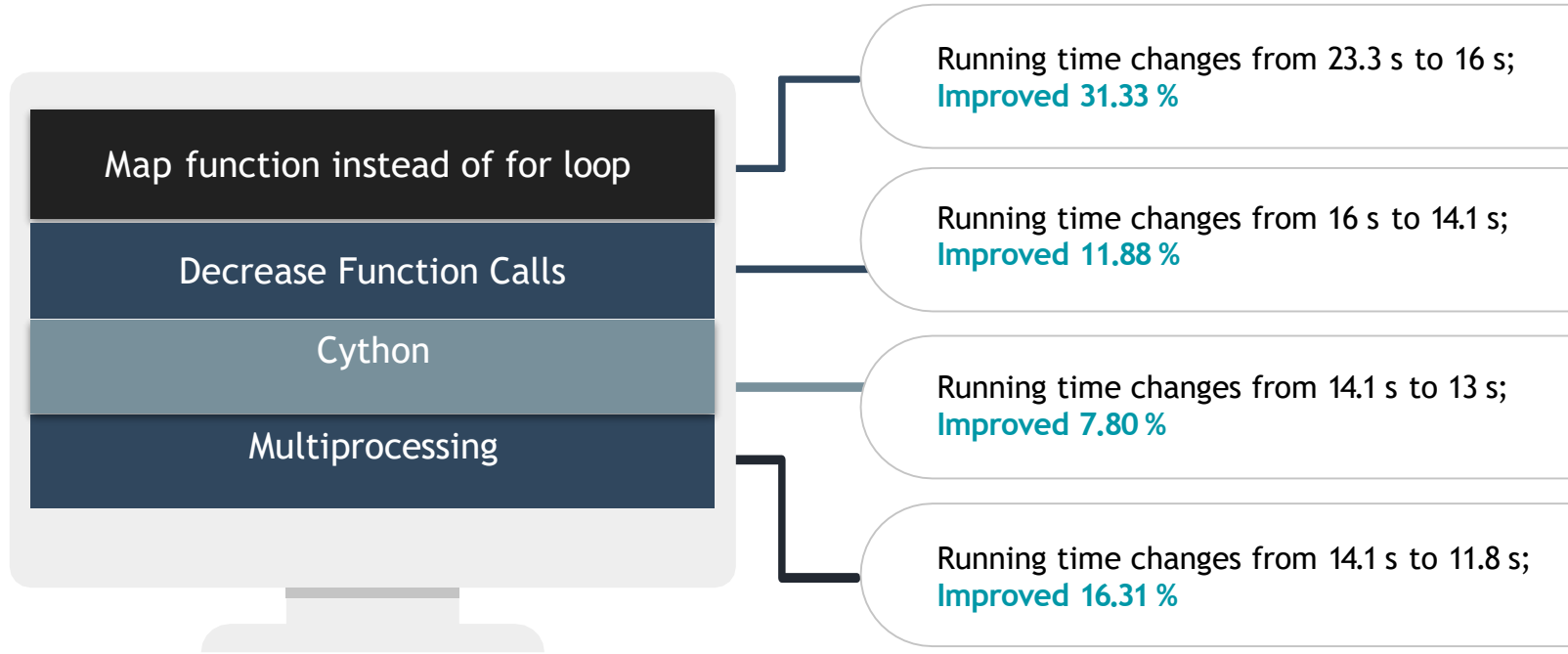
- 50% dropout rate to prevent overfitting
- GPU
- Batch processing

04

Model Performance

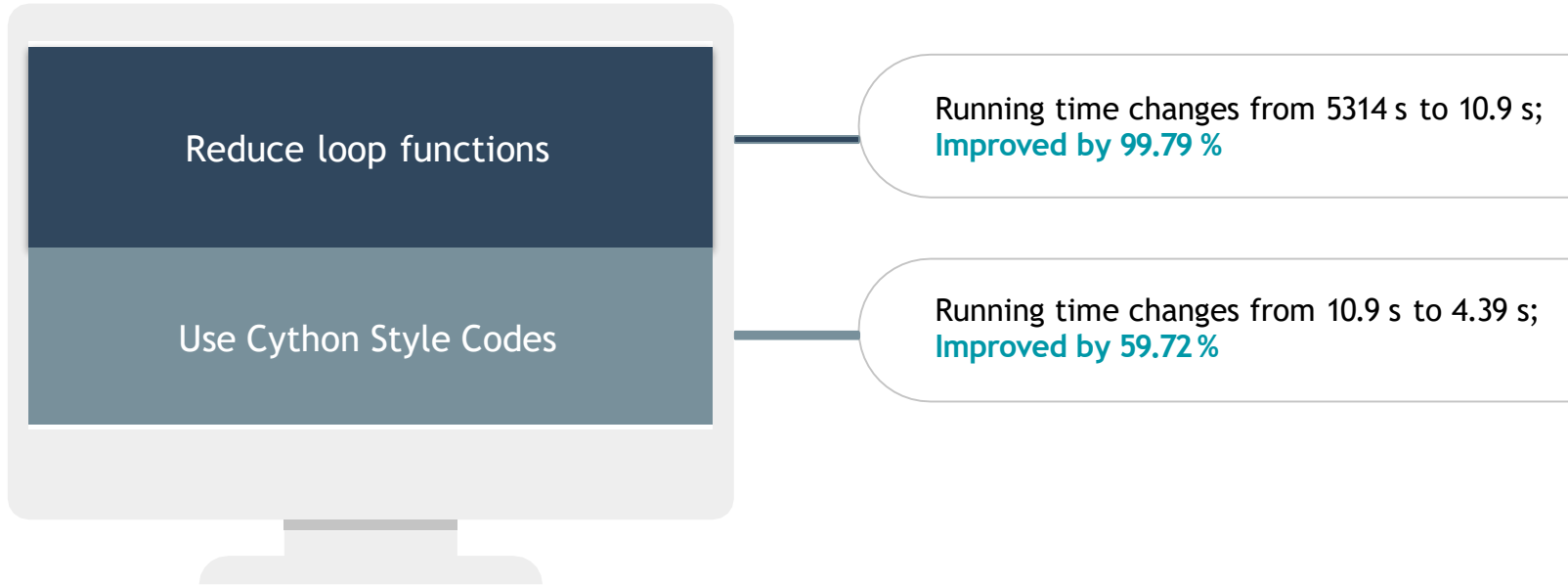
- Accuracy of 83.79%
- 10% improvement over the baseline
- top 5% among 1200+ participants

Optimization of Preprocessing



Overall, running time has improved 49.36 %

Optimization of SVM



Overall, SVM running time has improved 99.91%



Thank you!