

Final project - Proposal

Team information:

Team Name: **The West Coasters**

Team Members

- Marina Polupanova - marinap2 [Captain]
- Tirthankar Bhakta - tbhakta2
- Savita Manghnani - savitam2

Selected competition:

Option4 - Text Classification competition

State and goals of the project:

Our team has almost no experience with any of the neural networks and machine learning frameworks. For this program we intend to explore on following techniques and Machine Learning frameworks:

- **Random Forest**
- **Logistic Regression**
- **Support Vector Machine**
- **Recurrent Neural Network**

Machine Learning Frameworks

- **Tensorflow**

Tensorflow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries, and community resources that let researchers push the state-of-the-art in ML and developers easily build and deploy ML-powered applications.

The system is general enough to be applicable in a wide variety of other domains, as well. We will pick up one ML model in the TensorFlow library for our classification task.

- **Keras**

The Keras neural networks library supports both convolutional and recurrent networks that are capable of running on either TensorFlow or Theano. Keras deep learning framework was built to provide a simplistic interface for quick prototyping by constructing active neural networks that can work with TensorFlow.

In a nutshell, [Keras](#) is lightweight, easy-to-use, and has a minimalist approach. These are the very reasons as to why Keras is a part of TensorFlow's core API. The primary usage of Keras is in classification, text generation, and summarization, tagging, translation along with speech recognition, and others.

- **Scikit-learn**

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions. We are planning to use the Scikit's robust set of algorithm for:

- Regression: Fitting logistic regression models. For Logistic regression we will as well try GLMnet library, and will submit the results of the library which will look more competitive.

- Decision Trees: Tree induction and pruning for both classification and regression tasks
- SVMs: for learning decision boundaries

Programming language:

- Python