Northeastern University
EECE5644 – Pattern Recognition and Machine Learning
April 2018, Asha Chen-Phang, Emily Dutile, Nate Otenti, Tristan Sweeney

---

Classification in High-Resolution Brain Scans Using Apache Spark: Project Work Distribution

Asha Chen-Phang:

- Aided data analysis in Jupyter Notebook (python) on subsets of the dataset
- Aided with parameter tuning within the model code of the Spark/Scala program
- Worked on configuration to run a python script for preprocessing to be completed on EMR
- Prepared, wrote and edited the report and the presentation
- Wrote the Readme with instructions on how to build and run locally as well as run on EMR

Emily Dutile:

- Implemented and configured Makefile needed for standalone and AWS runs for Apache Spark
- Executed AWS/EMR Runs and gathered results for analysis
- Created Random Forest Model Training in scala using MLlib
- Data analysis in Jupyter Notebook (python) on subsets of the dataset
- Implemented Random Forest Prediction Phase in scala using MLlib
- Wrote Naive Bayes in scala with MLlib for testing/experimental purposes
- Feature Extraction and preprocessing in Jupyter Notebook using matplot lib, numpy, and pandas
- Report and presentation contributions

Nate Otenti:

- Model Comparison Tests in python (Jupyter notebooks) using pandas, numpy, and matplotlib
- Apache Maven integration (pom file) needed for building the project and running on EMR
- Helped with feature Extraction and preprocessing in Jupyter Notebook using matplot lib, numpy, and pandas
- Attempted AWS/EMR Runs
- Presentation and report contributions

Tristan Sweeney:

- Random Forest in python (Jupyter Notebook) using sklearn
- KNN in python (Jupyter Notebook) using sklearn
- SVM in python (Jupyter Notebook) using sklearn
- Executed numerous runs (locally) with entire dataset and pre-processed dataset for analysis
- Implemented Evaluations for models in Jupyter Notebooks using sklearn, matplotlib and numpy
- Executed AWS/EMR Runs
- Report and presentation contributions