CIS 6930: Trustworthy Machine Learning
Project Proposal: Title of the Project

Tanvi Jain
(Point of Contact)
tjain@ufl.edu

Nimish Bajaj
nimishbajaj@ufl.edu

September 21, 2021

1 Introduction

Determine whether explainable / interpretable ML techniques like LIME and others can provide useful insights into the cause of the unfairness.

A model's accuracy is not always enough to state weather it will perform well in the wild because the accuracy highly depends the data that was used to train and test the model. Explaining the results of a model, and identifying what lead to a particular classification can help provide insights into the model. Once you add interpretability to the model, its result can be easily understood by a domain expert can be verified for correctness. This greatly enhances the trustworthyness of the model.

We are proposing to build a text classification model and explaining the results using LIME, and validate if LIME is able to provide stable and meaningful features.

2 Background and Related Work

Since the data contains technical words which are sometimes outside of the english vocabulary, we are going to tune word embeddings for the problem domain.

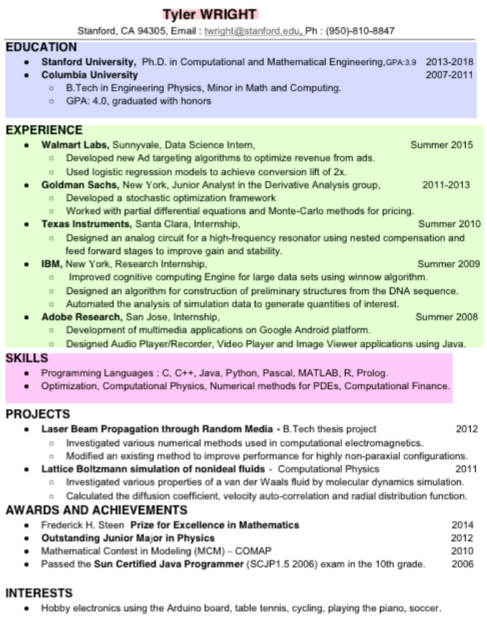
Write here.

You can cite related publications like so [1] and non-publications (e.g., websites) like so[1].

3 Proposed Approach & Plan

Following is the approach we are going to take:

1. Fetch data and take samples as per the distribution of data for train, test and validation
2. Clean and vectorize the data for Neural Network (NN) approach
3. Build a NN architecture to learn classification labels
4. Train, test and validate the neural network
5. Use LIME for model inferencing
6. Manually validate LIME output



Tyler WRIGHT

Stanford, CA 94305, Email : twright@stanford.edu, Ph : (950)-810-8847

EDUCATION

- **Stanford University,** Ph.D. in Computational and Mathematical Engineering,GPA:3.9 2013-2018
- **Columbia University** 2007-2011
  - B.Tech in Engineering Physics, Minor in Math and Computing.
  - GPA: 4.0, graduated with honors

EXPERIENCE

- **Walmart Labs,** Sunnyvale, Data Science Intern, Summer 2015
  - Developed new Ad targeting algorithms to optimize revenue from ads.
  - Used logistic regression models to achieve conversion lift of 2x.
- **Goldman Sachs,** New York, Junior Analyst in the Derivative Analysis group, 2011-2013
  - Developed a stochastic optimization framework
  - Worked with partial differential equations and Monte-Carlo methods for pricing.
- **Texas Instruments,** Santa Clara, Internship, Summer 2010
  - Designed an analog circuit for a high-frequency resonator using nested compensation and feed forward stages to improve gain and stability.
- **IBM,** New York, Research Internship, Summer 2009
  - Improved cognitive computing Engine for large data sets using winnow algorithm.
  - Designed an algorithm for construction of preliminary structures from the DNA sequence.
  - Automated the analysis of simulation data to generate quantities of interest.
- **Adobe Research,** San Jose, Internship, Summer 2008
  - Development of multimedia applications on Google Android platform.
  - Designed Audio Player/Recorder, Video Player and Image Viewer applications using Java.

SKILLS

- Programming Languages : C, C++, Java, Python, Pascal, MATLAB, R, Prolog.
- Optimization, Computational Physics, Numerical methods for PDEs, Computational Finance.

PROJECTS

- **Laser Beam Propagation through Random Media** - B.Tech thesis project 2012
  - Investigated various numerical methods used in computational electromagnetics.
  - Modified an existing method to improve performance for highly non-paraxial configurations.
- **Lattice Boltzmann simulation of nonideal fluids** - Computational Physics 2011
  - Investigated various properties of a van der Waals fluid by molecular dynamics simulation.
  - Calculated the diffusion coefficient, velocity auto-correlation and radial distribution function.

AWARDS AND ACHIEVEMENTS

- Frederick H. Steen **Prize for Excellence in Mathematics** 2014
- **Outstanding Junior Major in Physics** 2012
- Mathematical Contest in Modeling (MCM) – COMAP 2010
- Passed the **Sun Certified Java Programmer** (SCJP1.5 2006) exam in the 10th grade. 2006

INTERESTS

- Hobby electronics using the Arduino board, table tennis, cycling, playing the piano, soccer.

(a) Example Input

(b) Model prediction

Figure 1: Model output description

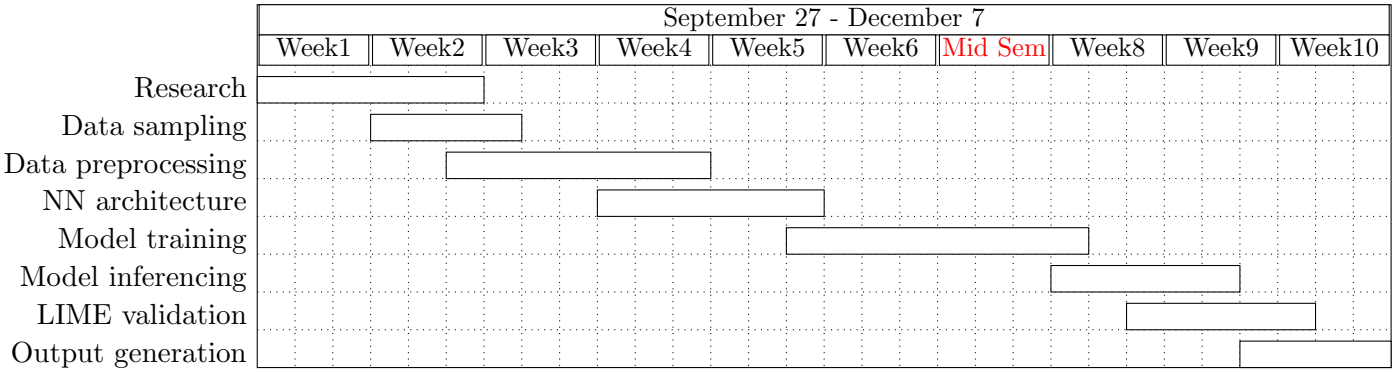


Figure 2: Gantt Chart

4 Timeline

References

[1] Vladimir Vapnik, Esther Levin, and Yann Le Cun. Measuring the vc-dimension of a learning machine. *Neural computation*, 6(5):851–876, 1994.

[1]Google scholar: https://scholar.google.com