Computer Science Capstone Topic Approval Form

The purpose of this document is to help you clearly explain your capstone topic, project scope, and timeline. Identify each of these areas so that you will have a complete and realistic overview of your project. Your course instructor cannot sign off on your project topic without this information.

Note: You must fill out and submit this form. The space beneath each number will expand as needed.

Any cost associated with application development will be the student's responsibility.

INFORM INSTRUCTOR:

Potential use of human subjects: (Y/**N**)

Possible use of proprietary company information: (Y/**N**)

ANALYSIS:

1. Project topic AND description:

   My project topic is an exercise in Natural Language Processing (NLP), a subset of the field of artificial intelligence. The exercise I'd like to do is sentiment analysis. I intend to analyze words in a sentence to determine the emotion behind the sentence. This can be either good or bad. I will achieve this by using deep learning and neural networks to create a program that teaches itself the sentiment of sentences it has never seen before to a high degree of accuracy. The base chance of a program to determine whether a combination of words is either good or bad would be 50%. My project will have a chance of success of at least 75%, nearing 90-99% accuracy.

   Client:

   The client for this project is a fictitious company called Data Solutions Inc. This company sells data, analytics, and tools to other businesses that look to make more informed decisions or receive guidance on complicated data-based issues. My role will be to work for Data Solutions Inc. and create a sentiment analysis tool that can be sold or reused in other analytics.

   Data Solutions Inc. and businesses who work with them will benefit from the creation of this tool and the completion of this project due to the new insight gained from the tool. With the use of neural networks, the program will understand what "good" and "bad" are. The program will be substantially faster than any human reviewer to tell the user whether a group of sentences is good or bad within the context of the document. This means a company could now objectively see if a product is being received well with positive customer reviews without using sales analytics or any other tool. The forecasting nature of the proposed program would allow the user to have a much more rapid response to subjective sentences; the applications for this type of program are endless.

2. Project purpose/goals:

   The proposed project's purpose is to show my ability to create a sentiment signal. Sentiment signals are widely used in the business world for many purposes. A few examples include business planning based on customer opinions, determining opinions and reception of goods/services, prioritization of customer services issues, and overall providing a greater understanding of what decisions should be made by a business. This sentence sentiment analysis tool's purpose is to showcase my capability to design a part of a system that inputs data and outputs reliable, accurate business forecasting information.

   Descriptive method:

   My proposed project will use K-means clustering as a descriptive technique to learn from the input data. The program recognizes and groups similar words, phrases, and sentences as training is executed.

   Prescriptive method:

   I intend to use a long-short-term memory (LSTM) recurrent neural network (RNN) to handle some of the most common issues with memory and attention in teaching a machine any language. Due to the nature of the neural network, inputs are

sent through the nodes to reach a mathematical calculation to a degree of accuracy. These outputs are then recycled as the inputs to another layer of input nodes (deep learning) which dramatically increases the model's accuracy. In a high-level description, the model is learning from past information to create new knowledge to learn from until there is an understanding of the words in sentences. This will allow the program to train on sentences and then, using these trained rules, see a sentence it has never before seen and predict its sentiment to a high degree of accuracy.

DESIGN and DEVELOPMENT:

1. Computer science application type (select one):
    - Mobile (indicate Apple or Android)
    - Web
    - **Stand Alone**

2. Programming/development language(s) you will use:
   Python, data analytics tools, libraries on top of Python (Numpy, Matplotlib, NLTK, etc.), Tensorflow, and Keras. I will write the core algorithm in Jupyter Notebook, with additional text editing in Visual Studio Code as needed. All are open-source.
3. Operating System(s)/Platform(s) you will use: Windows 10
4. Database Management System you will use: N/A
5. Estimated number of hours for the following:
    i. Planning and Design: 40
    ii. Development: 80
    iii. Documentation: 40 iv. Total: 160

6. Projected completion date: November 1, 2019

IMPLEMENTATION and EVALUATION:

1. Describe how you will approach the execution of your project:
    a. Data to meet my goal must be acquired.
    b. Clean and preprocess data.
    c. Create a model with the data.
    d. Train the model.
    e. Evaluate.
    f. If results are insufficient, preprocess data differently and return to step c.
    g. Document and create tools for visualization (graphs, charts, write-up)

STUDENT SIGNATURE

X. Joe Student

_____

By signing and submitting this form, you acknowledge that any cost associated with the development and execution of the application will be your (the student's) responsibility.

COURSE INSTRUCTOR'S NAME: A. M. Turing

_____

COURSE INSTRUCTOR APPROVAL DATE: 8/1/2024

Project Compliance with IRB Y/N: Y