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. Space beneath each number will expand as needed.

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

INFORM INSTRUCTOR:

Potential use of human subjects: (Y/N) No

Potential use of proprietary company information: (Y/N) No

ANALYSIS:

1. Project topic AND description:
   TOPIC: Identification/Classification of Major League Baseball Pitch Types
   DESCRIPTION: According to MLB.com, there are currently 13 varieties of pitches thrown in modern gameplay. During competition - lacking the ability to directly verify with the pitcher - it can be difficult for spectators, announcers, or competitors to identify one type of pitch from another. In some instances, the line between pitches has become blurred such that practitioners have invented new terms to describe them. For instance, in modern parlance, the "slurve" has become increasingly common - a pitch that falls somewhere between a slider and a curve. Is it possible to train a machine learning model to successfully identify pitches to a level comparable to that described by Major League Baseball's own system?

2. Project purpose/goals:

   The goal of this project is to develop a machine learning model capable of identifying what type of pitch was thrown, given the associated Statcast metrics.

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: Numpy, Pandas, Matplotlib, Jupyter Notebooks, SciKit-Learn
   SQL

3. Operating System(s)/Platform(s) you will use:
   Mac OSX 10.13.6 High Sierra

4. Database Management System you will use:
   MySQL 8.0.19

5. Estimated number of hours for the following:
   i. Planning and Design: 20
   ii. Development: 60
   iii. Documentation: 20
   iv. Total: 100

6. Projected completion date: April 25, 2020

IMPLEMENTATION and EVALUATION:

1. Describe how you will approach the execution of your project:

   Phase 1: Initial Data Analysis

   - Examine available data from seasons 2015-2018
   - Identify any rules changes or implementation changes that may affect homogeneity of the data set
   - Verify quality of the data within the data set
   - Verify quality and reliability of measurements
   - Handle any necessary imputation or transformation
   - Assemble data into one aggregated data set for analysis

   Phase 2: Exploratory Data Analysis

   - Search for features exhibiting strong correlation with pitch type
   - Identify features within the data set potentially useful to pitch type classification
   - Prototype machine learning models capable of pitch type classification

   Phase 3: Product Development

   - Refine machine learning models
   - Develop minimally viable product
     - End-User Interface
     - Analyze/Classify User data

   Phase 4: Documentation

   - Prepare data product for presentation
   - Integrate output from graphical EDA techniques

STUDENT SIGNATURE

_____

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

COURSE INSTRUCTOR'S NAME:

_____

COURSE INSTRUCTOR APPROVAL DATE:

Project Compliance with IRB Y/N: