Introduction

In this document you will find a small take home assignment, that will help us judge your programming skills and proficiency in python.
The objective of the assignment is for you to demonstrate your coding abilities along with basic understanding of machine learning and software development. Please note there is no one right solution. What matters is that you are able to solve the problem and able to explain your thought process behind it.

Tasks

The assignment consists of two small tasks

Cricket Prediction Trainer

For the first task, you will have to create a simple binary classifier, which is able to predict if the team chasing in a cricket T20 match, will win the match or not.
You have been provided with a dataset called **cricket_dataset.csv** which has the following schema

column name	**Description**
total_runs	Total runs scored till that ball
wickets	Number of wickets that have fallen
target	The target they are trying to achieve
Balls left	How many balls are left in the match
won	1 - if team chasing won, 0 if they lost. This is the field we want to predict with our model

Using the above provided data, your task is to train a model that is able to predict if the team batting second will win the match or not given **total_runs, wickets, target** and **balls_left** as input features.
Once the model is built, please evaluate it and document your evaluation of the model.

Note: We are looking for a very basic implementation, we do not expect the model to have very high accuracy, the purpose of this task is to judge your basic understanding of machine learning and how you go about solving a problem. Please document the model you decided to train, and why did you select the perticular model. Also any design decisions that you took.

Prediction API

For the second task, you will build an API with a single end point, that takes path to a CSV as input and then run prediction on the values in that CSV based on the model generated in task 1. The results of this will then be stored to disk in a CSV, and the API will return the path to this results csv
For this purpose we have provided you with a test csv called **cricket_dataset_test.csv**.

The CSV contains the following columns

column name	**Description**
total_runs	Total runs scored till that ball
wickets	Number of wickets that have fallen
target	The target they are trying to achieve
Balls left	How many balls are left in the match

When your API gets the input data, it will only run prediction on rows of the CSV that meet the following criteria

- Balls_left is less than 60
- Target is greater than 120

Once it has ran prediction on the rows, it will store the results in a result csv (value of won set to 0 or 1) and then store it on disk. The path to this csv will be returned by the API

Evaluation

The purpose of this assignment is to see your familiarity with Machine Learning and Programming in General. Your tasks will be evaluated on the following criteria

- Quality of code
- Documentation of the model and its evaluation
- Code Structure
- Performance of the model

As stated above, the performance and accuracy of the model is not as important as the documentation, your over all code quality and structure.

Best of luck, If you have any questions regarding the assignment, feel free to reach us via Email.