10/27/2020

Roshandeep Singh Saini

100766638

STATEMENT OF WORK-1002

MARCOS BITTENCOURT

Contents

[1.1 Executive Summary 2](#_Toc54693212)

[1.2 Problem Statement 2](#_Toc54693213)

[1.3 Analytics Rationale Statement 2](#_Toc54693214)

[1.4 Data 2](#_Toc54693215)

[1.5 Data Analysis Approach 2](#_Toc54693216)

[1.6 Project Plan 3](#_Toc54693217)

## Executive Summary

An analysis is conducted to predict whether a given dish is vegetarian or not, depending on the ingredients used in the dish. In simpler terms it is expected that that a dish with same ingredients, would result in a similar dish. Predictive modelling techniques would be used to find a solution which best classifies the dish as vegetarian or non-vegetarian in terms of diet.

## 1.2 Problem Statement

The problem statement is “To classify a dish as vegetarian and non-vegetarian based on the ingredients”.

Based on the ingredients used to prepare an Indian dish, it can be classified as a part of a vegetarian diet or a non-vegetarian diet. To gain insights on how dishes with similar ingredients can be classified similarly or differently based on region, flavor, and course.

## 1.3 Analytics Rationale Statement

The rationale of the analysis is to accurately predict whether a dish can be called a vegetarian dish or a non-vegetarian dish depending on the contents, flavor profile, or the course of meal. Dishes with similar ingredients can be profiled as a different course of meal as per Indian cuisine. Hence, the aim is to quantify the similarity and disparity.

## 1.4 Data

The data has been acquired from Kaggle open datasets. It is a raw dataset named “indian\_food”, which represents entirety of the testing data for August 5th to October 5th, 2020. It contains details about Indian food, the features used to classify them.

It has 255 records. These variables are: -

* Independent Variables- Actual measurement parameters of an Indian Dish which are name, ingredients, prep\_time, cook\_time, flavor\_profile, course, state, region.
* Dependent Variable- Classification of dish as part of the diet(vegetarian and non-vegetarian).

## 1.5 Data Analysis Approach

The methodology used will be to develop a classification model using machine learning approach to segregate Indian dishes into “vegetarian” and “non-vegetarian”. It is a scenario of Supervised Learning. This would require classification algorithms SVM, Decision Tree and an advanced modelling through TensorFlow(Keras) etc. depending on the complexity and the type of data received. This would be determined in the EDA and modelling phase of the project. The software tools used will be: -

* Python – for EDA (Exploratory Data Analysis), Data Cleaning, Model Building and Testing
* Jupyter Notebook – IDE for development
* Tableau/Power BI – for detailed visualizations

## 1.6 Project Plan

The table below contains the project tasks and their estimated completion dates.

|  |  |  |
| --- | --- | --- |
| **Phase** | **Task** | **Delivery Date** |
| **Project Organization** | Setup | October 28, 2020 |
| **Business Understanding and Problem Discovery** | Statement of Work(V1) | October 28, 2020 |
| **Data Acquisition and Understanding** | Data Acquisition and Understanding, Statement of Work(V2) | November 23, 2020 |
| **ML Modelling and Evaluation** | Advanced Predictive Modelling | November 23, 2020 |
| **Delivery and Acceptance** | Deployment of Software Pipeline | December 18, 2020 |