

CLOUD TECHNOLOGIES

CA675

Assignment-2 Cloud Application

Group-0 Mid-Way Report

We declare that this material, which we now submit for assessment, is entirely our own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. We understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should we engage in plagiarism, collusion or copying. We have read and understood the Assignment Regulations set out in the module documentation. We have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by us or any other person for assessment on this or any other course of study.

We have read and understood the referencing guidelines found recommended in the assignment guidelines.

Name	Student Id	E-Mail id
Aditya Kumar Yadav	18210034	Aditya.yadav3@mail.dcu.ie
Alekhya	18210618	Alekhya.singh7@mail.dcu.ie
Kashyap Krishnamurthy	18210248	Kashyap.krishnamurthy2@mail.dcu.ie
Paritosh Gupta	18210686	Paritosh.gupta3@mail.dcu.ie

Date: 11-03-2019

Productive Reviews Affirmation App

Introduction

Amazon is the world's largest e-commerce marketplace. An integral part of a product listed on Amazon is the 'Review'. Reviews, primarily given by people who have purchased a given product, forms a crucial part in the decision making process for a second person intending on purchasing the said product. A review strongly influences one to either purchase the product or decide against it. The Productive Reviews Affirmation App(PRAA) focuses on enriching the Review writing process.

Use Case Definition

The proposed solution is a Machine Learning(ML) powered App that can either be integrated with an existing website or deployed independently. The App primarily consists of a 'textbox' where a review for a given product can be entered; on clicking a submit button – the system processes the review and indicates if the review would prove to be 'helpful' or not. The system can be compared to a password-strength indicator where the person is presented with information as to whether the password is strong or not – should the person be interested he/she can update the password to a stronger one. Similarly, the PRAA system presents the review composer with information on the usefulness of the review – should he/she decide to action upon a poor review is optional.

Use Case Realisation

This section details the data being used and the architecture outline of the proposed solution.

Data Definition

The dataset chosen for the solution consists of reviews for fine-foods from Amazon and other Amazon categories. The data, sourced from Kaggle, spans a period of more than 10 years. The data consists of over 560,000 reviews up to October 2012. Link to the dataset on Kaggle – [Click here](#).

Records include –

- | | |
|---------------------------------------|---|
| • <i>Id</i> | - Record Id |
| • <i>ProductIdUnique</i> | - identifier for the product |
| • <i>UserIdUnique</i> | - identifier for the user |
| • <i>ProfileNameProfile</i> | - name of the user |
| • <i>HelpfulnessNumeratorNumber</i> | - users who found the review helpful |
| • <i>HelpfulnessDenominatorNumber</i> | - users who indicated if the review was helpful |
| • <i>ScoreRating</i> | - between 1 and 5 |
| • <i>TimeTimestamp</i> | - time of the review |
| • <i>SummaryBrief</i> | - summary of the review |
| • <i>TextText</i> | - the review body |

Solution Outline

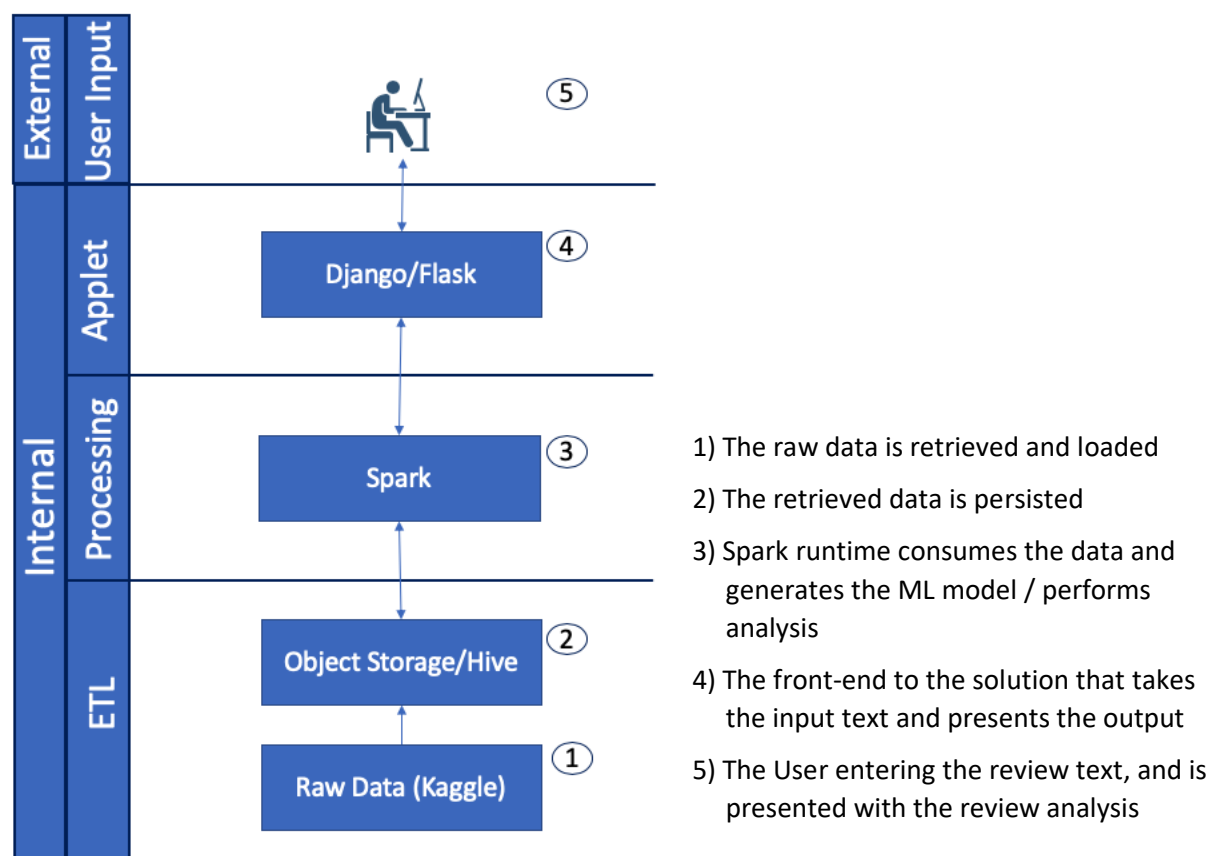
The proposed solution has two phases of operation –

- Training Phase
- Live Phase

Training Phase - The data sourced from Kaggle will be persisted in a chosen data store, Hive or an Object Oriented Storage such as S3 are the platforms being considered. The persisted data will then be processed in a Spark runtime; the data will be wrangled, modelled and fitted into a chosen ML Algorithm. The goal of the trained ML model would be to analyse new text(reviews) and ascertain if the review is useful or not based on the training data.

Live Phase - The Applet front-end, built with either Django or Flask, takes in the user input and passes it onto the 'Processing' phase. The data will be processed and the review-usefulness analysis will be presented to the end user.

Solution Outline



Excepted Contributions

Team members will contribute in all the areas of the solution and will monitor each stages of the pipeline. The table presents the member who would shoulders a given platforms functionality.

Pipeline	Primary Responsibility
Raw Data Ingest – Hive/Object Storage	Aditya, Kashyap
Processing Engine- Spark (PySpark)	Kashyap, Alekhya
User Interface	Alekhya, Paritosh
Integration Services	Aditya, Paritosh