

Principles of Big Data Management

COMP-SCI 5540 [Fall Semester 2024]

Research Project (40 Marks)

Submission Dates: Oct 29th, 2024 (Tue) Phase-1 (5 Points)

Submission Dates: Nov 15th, 2024 (Fri) Complete (30 Points)

Viva as per the schedule: (5 Points)

**Instructions:**

* Research project is a group-based activity.
* Submit your documentation as a single document, and the code as separate file or link.
* Late and copied work won’t be graded and will result as ZERO credit.

**Project Title: Resume Evaluation System using NLP**

**Group Members:**

|  |  |
| --- | --- |
| **Name** | **Student-ID** |
|  |  |
|  |  |
|  |  |
|  |  |

**Abstract**

Selection of suitable candidates for a job is critical / crucial for the companies. Due to varied recruitment and multi-skilled job descriptions, traditional hiring methods have becoming insufficient. Conventional techniques include, manual searching of applicants, reviewing their resumes, shortlisting and calling for interviews. Some of the process have been automated already but still this process need be smarter and more trustworthy. Resumes / CVs that are not always structured so they need to be processed to yield valuable outcomes and provide a ranking list according to the preference and requirement of the company.

1. **Introduction**
2. **Project Scope**

* Converting unstructured data to structured format using NLP
* Extracting the relevant information from the resume and assigning them identifier values
* Ranking of resumes

1. **Project Limitations & Constraints**
2. **Feasibility Study [Technical | Operational]**
3. **System Requirement Specifications (SRS) [MDRE | Bespoke]**
4. **System Design**
5. **Use-Case Diagram**

Use-case diagram is a graphical representation of user's possible interactions with a system. It shows functionality (use-cases) and different types of users | stakeholders (actors) of the system. Use-cases are represented by either circles or ellipses and actors are depicted in form of human symbol.

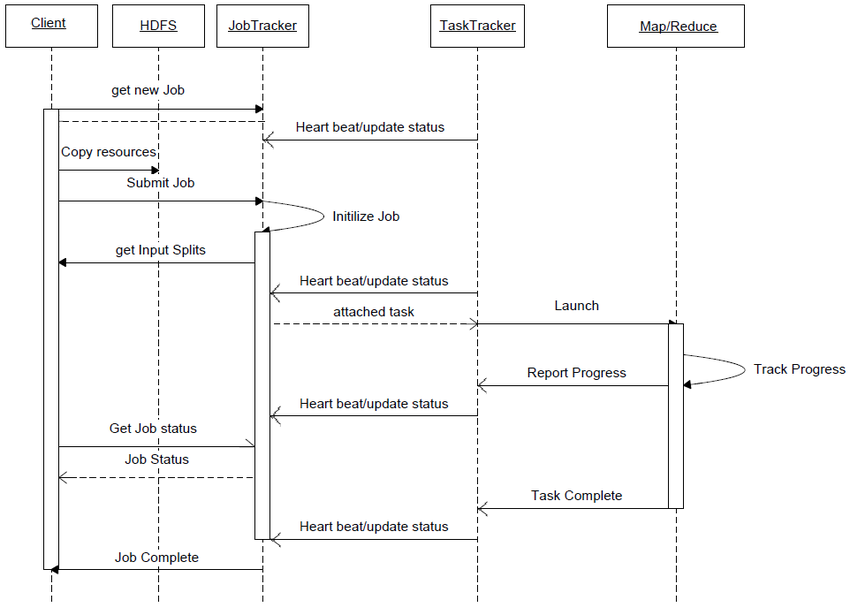
|  |
| --- |
| 10 Tips to Create Professional Use Case Diagram - ArchiMetric |
|  |

**Possible Actors in Data Science Projects:**

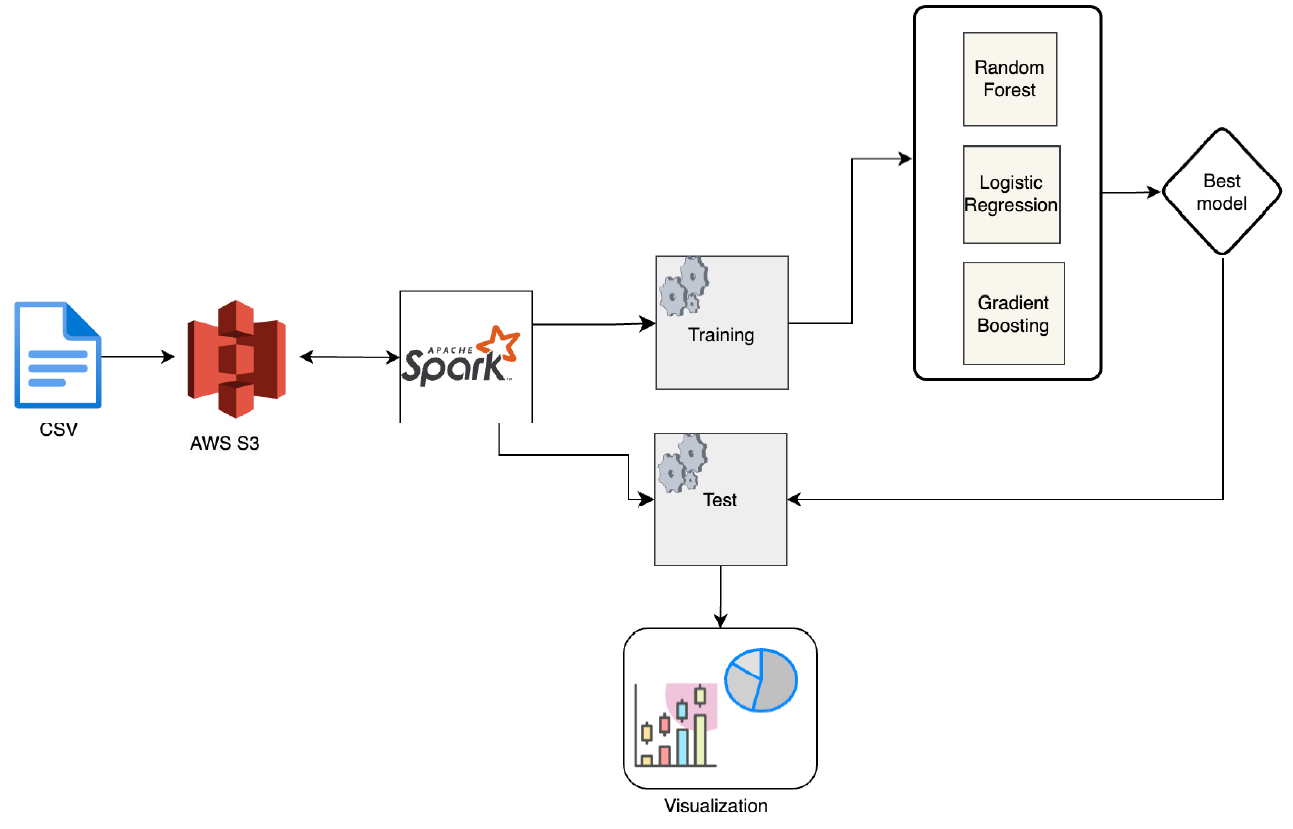
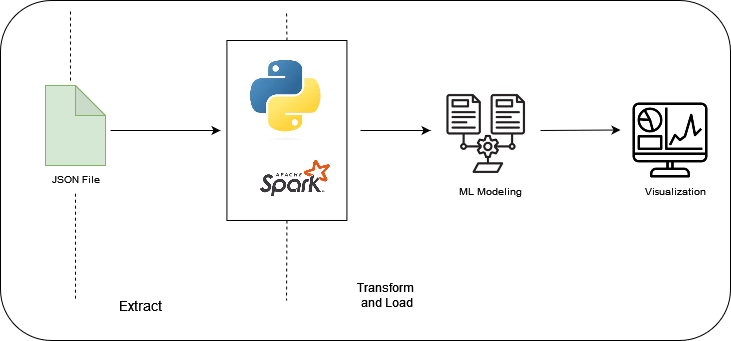
* ML and data professionals
* Software engineers
* Business stakeholders
* Product owners
* End users
* Security professionals
* Quality assurance professionals

1. **Sequence Diagram**

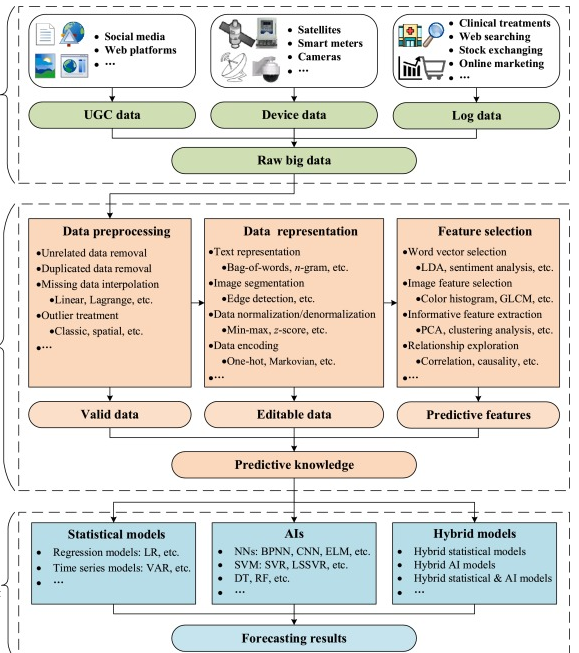
This is just exemplary figure, indeed you will be pasting the sequence diagram of your project.

****

1. **Data Design**
2. **ETL Process (Explanation)**



1. **Data Management**



* What data is being selected for this project?
* Why was this data selected (and other datasets excluded)?
* How was the data obtained?
* What are known issues in the data?
* What does the data look like? (Nominal, Ordinal, Discrete, Continuous)
* How did you alter the data (transformations, imputations, other data cleaning techniques applied, etc.)
* Where is the data located?
* How frequently is the data refreshed?
* Is selected data can be good input for yielding correct business plans?

1. **Data Analytics and Modeling**

* Descriptive | Diagnostic | Predictive
* Statistical or ML

1. **Data Visualization**

|  |  |
| --- | --- |
| data-visualization-project · GitHub Topics · GitHub |  |
|  | Stacking Machine Learning Models for Multivariate Time Series | by Alvin T.  Tan, Ph.D. | Towards Data Science |

1. **Code (Only main modules in report | submit code separately)**
2. **Test Cases**
3. **References**