ML based Career Service (MBCS)

INFO 7390 ADS Project Proposal

April 06, 2018

Team :Alpha Beta Gamma

Sreerag Mandakathil Sreenath
001838559
mandakathil.s@husky.neu.edu

Shreya Chudasama
001828562
chudasama.s@husky.neu.edu

Aahana Khajanchi
001824402
khajanchi.a@husky.neu.edu

Table of Contents

Project Proposal **3**

Stakeholders **3**

- The Supervising authority 3
- The Beneficiaries (or users) 3
- The Suppliers (or service providers) 3
- The Recruiters 4

The steps involved **4**

Pipeline **6**

- Luigi 6
- Docker 6
- Amazon S3 6

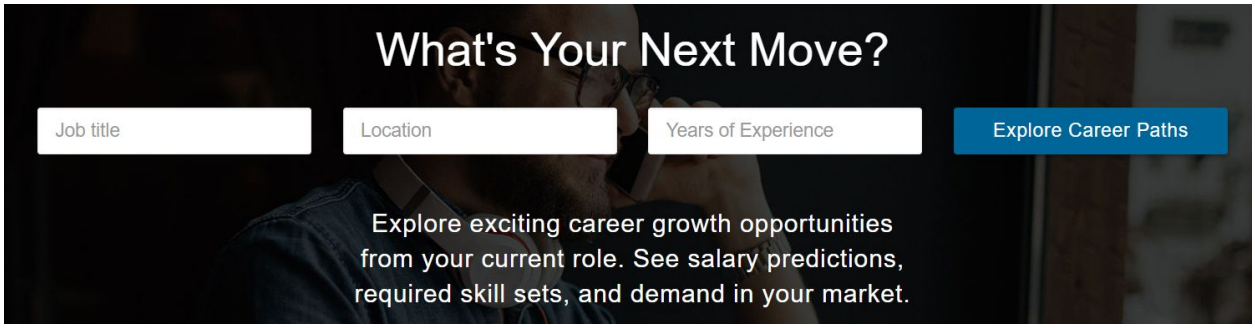
Advantages **6**

Milestones **7**

References **7**

Project Proposal

MBC is Web based platform for candidates who are seeking a job and recruiter who are looking for the right talent. The platform scrapes and analyzes publicly available job posting data from sites like Glassdoor, LinkedIn, etc. The system will recommend the candidate which job title will fit him/her based on their resume. The system will also give job links to the application form. The platform will also help us to gain statistical information about the latest job trend.



Stakeholders

The Supervising authority

The Supervising authority is a person that governs the entire system. Having supervisory powers over some aspects of management decision-making. Basically a person with more power than others.

The Beneficiaries (or users)

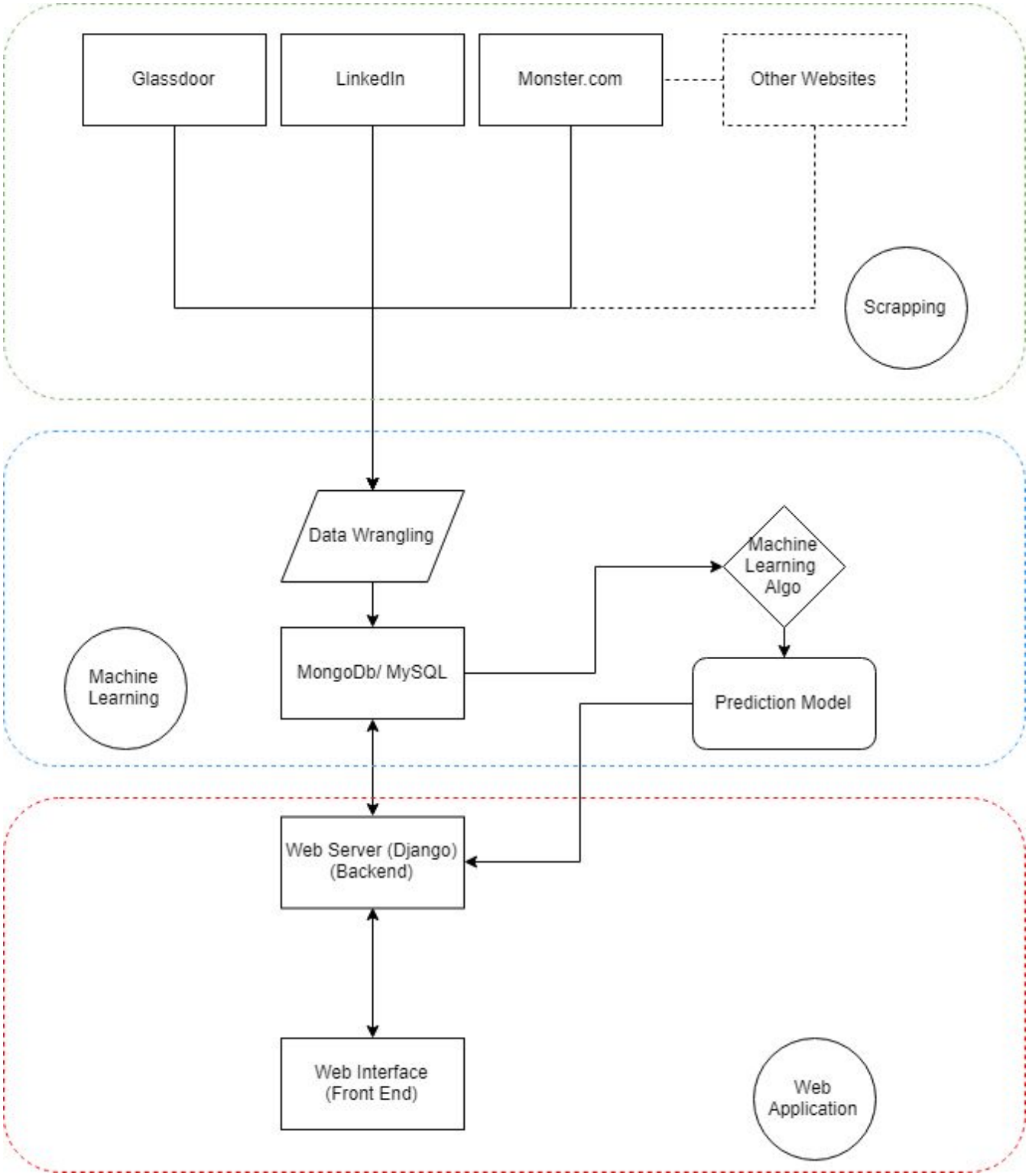
The beneficiaries (or users) are all natural persons or corporate body, students , recipients of the services provided by MBCS.

The Recruiters

A person whose job is to enlist or enroll people as employees as members of an organization.

The steps involved

1. Scrapping information and skill sets from websites such as Glassdoor
    a. This will be done by giving the url as input to a web scraper
    b. Library such as scrapy will be used to run multiple web crawlers on a single website to optimise scrapping process
2. Analysing the data and segregating the content to skills and tools mandatory and the expertise level
    a. The data will be stored in either a mongoDB or MySQL database for further processing
3. To try out different machine learning models and hybrid techniques to optimise the result
4. To make a web based application to interact with the system through which the user can upload his/her data
5. Parse the users resume and scrape relevant information i.e. tools, skills, location, work experience and serve it as parameters for our prediction model.


Glassdoor
LinkedIn
Monster.com
Other Websites
Scrapping
Data Wrangling
Machine Learning Algo
Machine Learning
MongoDb/ MySQL
Prediction Model
Web Server (Django) (Backend)
Web Interface (Front End)
Web Application

Deployment Details

1. Luigi

   Luigi is a Python package that helps you build complex pipelines of batch jobs. It handles dependency resolution, workflow management, visualization, handling failures, command line integration, and much more.

2. Docker

   Docker is an open source tool that automates the deployment of the application inside software container.

3. Amazon S3

   Amazon Simple Storage Service (Amazon S3) is a web service that provides highly scalable cloud storage. Amazon S3 provides easy to use object storage, with a simple web service interface to store and get any amount of data from anywhere on the web.

Advantages

1. Discover number of tech pros skilled in everything from the most sought-after skills like Android and Big Data to the most common like Java.
2. User just have to upload their resume and everything else would be handled by the system,user won't have to write down the skills.
3. Predict salary expectation in different profiles and which is best suitable for that resume.

Milestones

Timeframe	Delivery
Day 1	Write Scrapping code
Day 2-3	Data Preprocessing and Exploratory Data Analysis
Day 4-6	Model Building, Training, Selection
Day 7-8	Deployment of models on cloud and build web application
Day 9-10	System integration and documentation

References

- https://luigi.readthedocs.io/en/stable/
- http://boto3.readthedocs.io/en/latest/guide/s3-examples.html
- http://boto.cloudhackers.com/en/latest/s3_tut.html
- https://towardsdatascience.com/beginners-guide-to-data-science-python-docker-3181fd321a5c
- https://doc.scrapy.org/en/latest/intro/tutorial.html