iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XI **Month of publication:** November 2021

DOI: https://doi.org/10.22214/ijraset.2021.38911

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

Portfolio Builder and Career Recommender by Scraping Data Using Flask and Tensorflow

Krishangi Pandey[1], Chinta Aaron John[2], Deepak Nanjundan[3], R. Sri Saiomkar[4], C. Gangadhar Reddy[5]
[1, 2, 3, 4, 5]*Computer Science, Vellore Institute of Technology, Vellore*

Abstract: ***In this day and age, the use of physical copies for portfolios and resumes has largely been replaced with the use of online documents to showcase the skills and experience of the individual in question. However, this too is beginning to give way to a newform of a portfolio- a website that communicates all of the above information in a document, but in an aesthetic and elegant way. While this was only usually required or recommended for a specific portion of the workforce, an increasing number of employers have begun to prefer this creative approach, making it more appealing to any aspiring applicants. In this project, we plan to build a website that will take data from users and build a custom portfolio using a variety of pre-developed templates and designs. The user just needs to provide their LinkedIn profile which will then be used to scrape the data of the user, usin Flask, Selenium, and BeautifulSoup. After the information and details of the user has been extracted, our website will provide various templates to choose from. The user will then have the option to build the website or download the zip file of the created website which could be used in future or could be hosted by the user. Hence, in this paper we aim to build a website that would help create a resume quickly and professionally without the hassle and struggle of figuring out how to format everything or what fonts and styles to use. We will help generate a dynamic resume that will highlight the skills, achievements and talents and also be able to showcase the user's thoughts and personality. Moreover, using Naive Bayes, we will provide a career recommendation to the user, which will guide steer them towards the most suitable profession for their skills***
Keywords: Portfolio, Flask, Selenium, TensorFlow, Naive Bayes, Beautiful Soup, Machine Learning.

I. INTRODUCTION

A portfolio gives the employees a platform to showcase their achievements, skills, experiences, and characteristics. It gives a chance to the users to show their collection of records that reflect and highlights some samples of their best life's experiences, values, beliefs, and achievements. The information included in a portfolio, can impact one's ability, both as an individual and as a professional, to be a useful tool in marketing themselves to employers, corporations, colleges, and universities. A portfolio similar to a resume, can be indicative of the abilities and what an individual has to offer in their chosen field. It's safe to say that everyone can enjoy the benefits of having a website of their own, but for some career paths its almost a requirement. If an individual is a designer, a photographer, videographer, website designer, or a graphic designer, artist, web developer, or even have experience in the field of marketing and branding, a lack of an online portfolio will almost certainly hinder their application. The creativity with which this portfolio is created itself highlights the talents of the user. These days, a lot of people would prefer to enhance their online presence, create websites, in order to show their work, with a personal website. In addition, some of the advantages that we will discuss below, can be applied to a wide range of applications.

There are various benefits of having a portfolio website:

1) It provides a platform to showcase our works.
2) It makes us more accessible for the clients and companies;
3) It creates an online presence which further helps in reaching out to mote people.
4) Displays consistency and professionalism.
5) Reflects our personality and creativity.

Since having an online portfolio is important as established earlier. But creating a website is a very time taking process which needs to be invested to get a good online portfolio and most of the people don't have that much time to invest into it.Even if someone is willing to create a website and has time for the other obstacle that comes into picture is the technical skills.Creating a website needs us to have a good technical knowledge about the frontend as well as the backend development which is something not everyone has. Because of these issues we came up with the idea to create a website which could do the same within minutes.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

The context of a resume makes a dramatic difference on how a employer sees a prospective employee. The need for a resume builder that can keep in mind every small detail that can help persuade an employer to hire is important. The advertisement of the skills and achievements in a such a way that's It's an eye catcher for the interviewer. We need to build an application that simplifies the task of creating a resume for individuals. The systems should be flexible to use and reduce the need of thinking and designing an appropriate resume according to qualifications. The system will be developed to provide an easy user-friendly system for creating a professional portfolio resume.

II. MOTIVATION

The portfolio website is the need of the hour for career growth and development irrespective of the domain or job. A portfolio is a collection of work that shows your performance and learning; it is essential for every artist's development. Online portfolios are also beginning to branch beyond being a requirement merely for small set of careers. It showcases your personality to potential employers and organizations. It is a useful tool to include in an interview. It provides tangible proof of your skills and abilities and demonstrates through a pie chart what level you are qualified for that specific job to the employer. Our project is a method for not only easily creating an attractive portfolio as per user's requirements, but also helping others who wish to create a similar web-based portfolio by providing them with ample tools and templates to create it as per their liking and giving a platform for the employers to view and get suggestions based on the profile what fields the user is best suited for.

A. Social Reasons

Nowadays, almost everyone with an internet access is on a social media platform, or other similar websites. LinkedIn has become synonymous with most businesses, with millions of employees' profiles being visible for reference. Social media profiles offer a simple and concise way of showcasing information about a user. In keeping with this trend, using a website to display the skills, experiences, and projects of an aspiring applicant, appeals to employers of all ages, and also highlights the creativity of the user.

B. Technical Reasons

LinkedIn has almost 740 million active users and many of them are either from a non-computers background or doesn't have any knowledge on web development. With the use of the latest web development frameworks, we can ease the process of building portfolios for these by providing pre-designed templates. The provided templates will be more appealing because they're easily customizable and easy to read. Due to the flexible online portfolio, you can be able to show your personality by choosing the design, layout and content you write.

C. Environmental Reasons

As the number of job applications rise, with an increasing number of people looking for jobs, multiple copies of paper-based resumes waste a lot of paper due to their requirement to be printed. This problem can be solved by using web-based resumes, which have the additional advantage of being more easily accessible.

D. Economic Reasons

A web-based resume is an inexpensive and effective method to build highly attractive and customizable resumes. It also saves a lot of time by providing pre-designed templates. Many people can't build a website on their own and for getting it designed, they either have to pay a lot, or spend a lot of time learning the basics in order to create a rudimentary website. However, our website is completely free to use, giving it an advantage over existing implementations, and helping people in need to showcase them talents in an aesthetically pleasing manner.

III. PROPOSED WORK

From the frontend home page, we take the LinkedIn profile link that is to be scraped and make a portfolio. This link is passed to linkedin_scraper.py file where we initiate a testing browser for automation of the LinkedIn profile using selenium and chrome driver. Initially we login to LinkedIn website by using selenium. Now we use the link provided by the user and load their profile on the test browser. Now we click all the see more buttons on the profile for loading the entire website. Taking BeautifulSoup tool to full advantage we parse the HTML through soup and make it as a soup editable format. By using this soup, we will scrape required data from website by using class names, ids and the element type (e.g. Div, button, forms....). Now we convert entire data to JSON format. All the JSON data will be stored in the static folder.

©IJRASET: All Rights are Reserved

Next step is to place all the data into the templates we created. For this we use the FETCH function for getting the data from the JSON file which is in a static folder. Now we use a call back function after executing the fetch and extract data from the JSON file. Now we use javascript for dynamic creation of elements and add the data to these elements and place this data on the HTML DOM. We used the zip File function to zip all the files that are required to host the website and provide those files to the user.

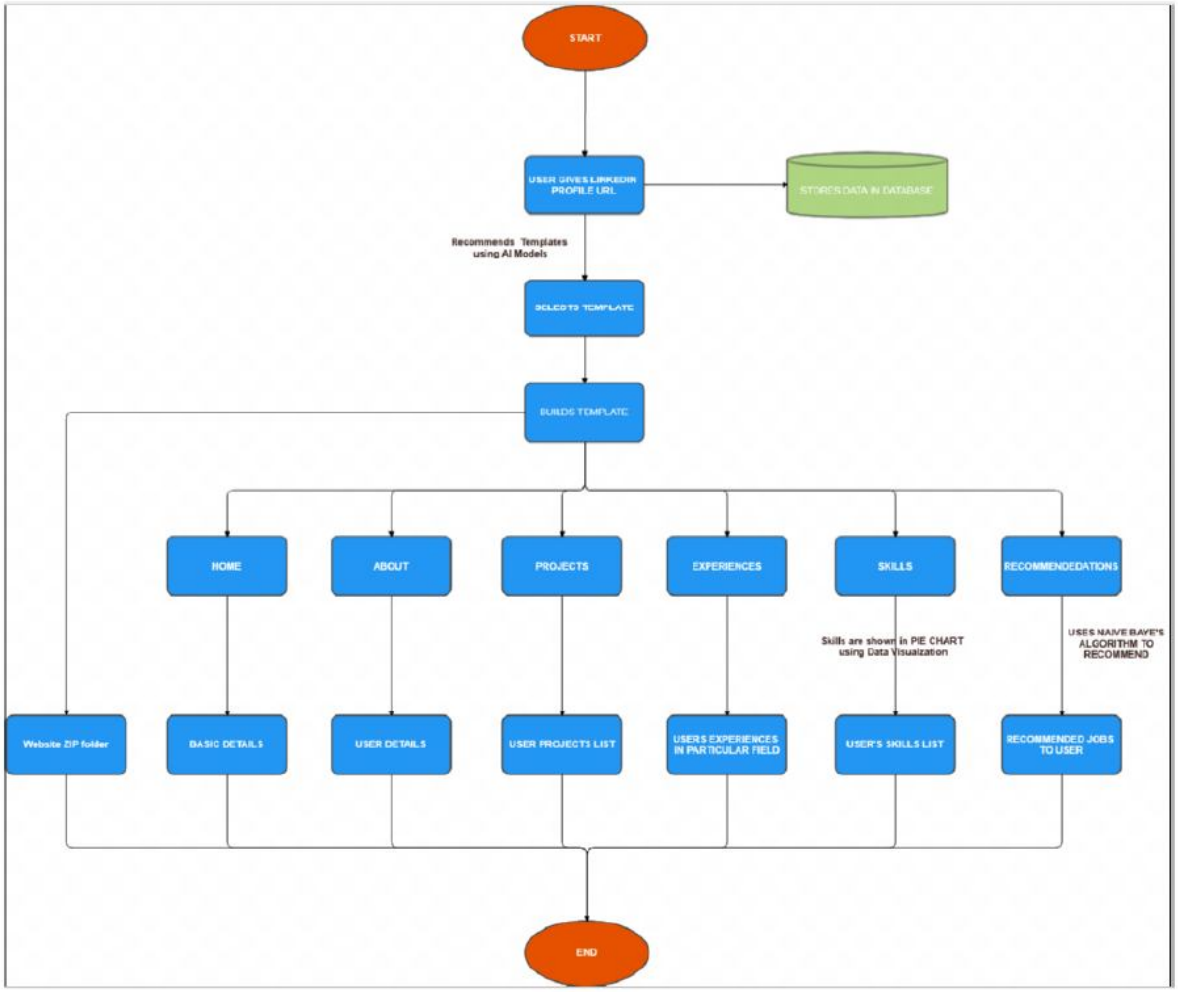


Fig. I Flowchart of our website

IV. THEORETICAL NOTIONS

We will build a website that will enable users to build a stylish resume online without any extra cost. The user will have to give their LinkedIn id and all the details of the user will be fetched from LinkedIn and a resume will be created that can be exported as pdf and a portfolio of the user will be created and a zip file of the same will be provided. We will also be providing a wide catalogue of templates for the user to choose from. We will be using web scraping to extract data from LinkedIn and convert the data to json format and use them at the backend. Moreover, we will also train a neural network to identify occupations based on the achievements present in the profile. We have used BeautifulSoup for extracting data from the web pages, and used “Selenium” for automating websites at places where we need to click “see more” and using flasks as backend.

A. API

An application programming interface is a computer interface that specifies how different software programmes or mixed hardware-software intermediaries communicate with one another. It specifies the types of calls or requests that can be made, how they should be made, the data formats that should be used, and the conventions that should be followed, among other things. It can also include extension mechanisms, allowing users to customise and expand existing features in a variety of ways. To ensure interoperability, an API may be completely custom, unique to a component, or based on an industry standard. APIs enable modular programming by

©IJRASET: All Rights are Reserved

hiding details, allowing users to use the interface regardless of the implementation. The word is currently most commonly used to refer to Web APIs. Programming languages and software libraries and computer operating systems, and computer hardware all have APIs. An API (application programming interface) simplifies application development by abstracting the underlying implementation and only revealing the artefacts or behaviour that the developer requires. While a graphical interface for an email client may provide a user with a button that performs all of the steps for fetching and highlighting new emails and API for file input/output may provide a feature that copies a file from one location to another without requiring the developer to understand the file system operations and taking place behind the scenes.

B. Web Scraping

Web scraping is a way of extracting vast volumes of data from websites in an automated manner. The majority of this data is unstructured HTML data that is translated to structured data in a spreadsheet or database before being used in different applications. To extract data from websites, web scraping can be done in a variety of ways. These involve using online platforms, specific APIs, or even writing your own web scraping code from the ground up. Many large websites, such as Google, Twitter, Facebook, LinkedIn, Stack Overflow, and others, provide APIs that let you access their data in a standardised format.This is the best choice, but there are other places that either don't allow users to access vast volumes of data in a standardised format or aren't technologically advanced enough. In that case, scraping the website for data using Web Scraping is the best option. The crawler and the scraper are needed for web scraping. The crawler is an artificial intelligence algorithm that searches the internet for specific data by following links across the network. A scraper, on the other hand, is a tool designed to remove information from a website. The scraper's design can vary greatly depending on the project's complexity and scope, but it must be able to extract data quickly and accurately. One might wonder how this works. Web scrapers will retrieve all of the data on a platform or only the data that a user is looking for. It's best if you define the details, you're looking for so that the web scraper just extracts the information quickly. For example, you may want to scrape an Amazon page for information on the different types of juicers available, but you just need the information on the different juicer models, not the customer feedback. When a web scraper needs to scrape a site and it is first given the URLs of the pages it needs to scrape. The HTML code for those sites is loaded and a more advanced scraper can extract all of the CSS and Javascript elements as well. The scraper then extracts the required data from the HTML code and outputs it in the user-specified format. The data is usually stored in the form of an Excel spreadsheet or a CSV file, but it can also be saved in other formats like a JSON file.

C. Beautiful Soup

BeautifulSoup is a library of Python that pulls data from HTML and XML files. It works with your favourite parser, so you can navigate, search and modify the parser tree idiomatically. It generally saves schedulers hours or working days. It creates a parsing tree to use for the extraction of HTML data, which will be useful for web scraping. Leonard Richardson was initiated by Beautiful Soup, a contributor to the project and is furthermore supported by the paid open-source maintenance subscription, Tidelift. Python 2.7 and Python 3 are available.

D. Selenium

Selenium is an open-source web browser automation tool. It has one interface that allows you to write test scripts in Ruby, Java, NodeJS, PHP, Perl, Python and C# programming languages, among others. A browser driver will then run these scripts on your device in a browser instance Web browsers are automatic with Selenium. The web app tests, which allow developers to ship up new releases quicker and more confidently, are most famous. Selenium's flexibility is part of why it is so popular. Anyone who codes on the web can use Selenium to test their code and application, from individual developers to debugging engineers who perform a rapid series of tests to visual regression tests.Today, most desktop/mobile browsers have integrated support for Selenium 11 automation testing. Consumer browser vendors such as Firefox, Chrome, Safari, IE and Opera are developing and shipping drivers to their browser. Others have contributed to the Selenium project since it was first open-sourced by including thirdparty drivers for specialised browsers such as BlackBerry 10 and HtmlUnit as well as integration bindings with developer boards such as PhantomJS,Qt, etc.

E. Flask

It is a micro web framework it written in Python. This is known as a microframe, as it requires no special resources or the library to use it. There is no database abstraction layer form the validation or any of the other components, which is dependent on the third-

©IJRASET: All Rights are Reserved

party libraries in order to accomplish a common task. Extensions, on the other hand, can also be used in order to allow functionality of whether they are in association with the Plunger in and of itself. The object relational map-add-ons to the existing form of technology, management, download,, several open authentication technologies, and some of the common tools that are related to the frame structure. Pinterest and Linkedin, you'll find both apps using the Flask framework. As it is, in the normal case, the equivalent of a Flask web application is, once again, of course, the Piston is considered Pythonic than the Django web framework. The flask is easy to get up and running as it is, because there is no boilerplate code in order to execute in a single app.

F. JSON

A web-based resume is an inexpensive and effective method to build highly attractive and customizable resumes. It also saves a lot of time by providing pre-designed templates. Many people can't build a website on their own and for getting it designed, they either have to pay a lot, or spend a lot of time learning the basics in order to create a rudimentary website. However, our website is completely free to use, giving it an advantage over existing implementations, and helping people in need to showcase them talents in an aesthetically pleasing manner.

G. TensorFlow

TensorFlow is an end-to-end open-source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries. It allows developers to create large-scale neural networks with many layers. For a web-based resume it provides a platform to use its tools and libraries for Classification, Perception, Understanding, Discovering, Prediction, and Creation.

H. R

R is a programming language used in web-based resumes to compute, collect, and analyze information. It has a quality of plots that the user can work efficiently while predicting and processing data. These include symbols in mathematics and other statistical formulae for occupational, behavioural prediction.

V. RELATED WORK

A. VERSHININA A.A., KISELEVA I.A., KOROTKOV A.V., NIKOLAEVA T.P., GORELOVA T.P. (2018). Market for corporate securities and building the CALM portfolio model

This paper simplifies the project portfolio selection process by developing a framework that separates the work into distinct stages. Each stage accomplishes a particular objective that creates inputs to the next stage. At the same time of users are free to choose the techniques they find the most suitable for each stage, ensuring that they feel as though they are in control.

B. Oleksandr Terentiev ,Tatyana Prosiankina-Zharova, Volodymyr Savastiyanov, Valerii Lakhno, Vira Kolmakova (2021). The Features of Building a Portfolio of Trading Strategies Using the SAS OPTMODEL Procedure.

This paper describes the original information technology of algorithmic trading, designed to solve the problem of forming the optimal portfolio of the trade strategies. The methodology of robust optimization and using the Ledoit–Wolf shrinkage method for obtaining stable estimates of the covariance matrix of algorithmic strategies and it was used for the formation of a portfolio of trade strategies. The corresponding software was implemented by SAS OPTMODEL Procedure, which is also discussed in-depth.

C. JN Mikhail (2021) Building Your Scholarship Portfolio: One Brick at a Time

This paper summarizes some of the research done to date on this subject and briefly discusses the need for and uses of an agreed portfolio categorization system, and proposes a first approach to establishing a number of broad categories based on the portfolio field or end results being produced by similar projects. It also describes a current project to develop such an agreed system with a global virtual team.

D. Sulaimon Olanrewaju Adebiyi, Oludayo Olatosimi Ogunbiyi, Bilqis Bolanle Amole (2021) Artificial intelligence model for building investment portfolio optimization mix using historical stock prices data.

This paper gives an idea of a sample of these methods and attempts to suggest an approach which builds on the strengths of existing AI-oriented methods to develop an integrated strategy that can be used to support managers in making critical decisions and concerning project portfolio optimization. The strategy allows flexibility in the selection of the methods to suit the culture and environment of the particular organization.

©IJRASET: All Rights are Reserved

E. K Savenkov (2020) Building Multi-Purpose MT Portfolio.
This paper makes the claim that, in order to provide the best value to the organization, the portfolio must contain a balance of project types and risk levels the number of projects must be limited to ensure that all projects can be resourced effectively.

F. LX Wang, J Ho-Shek, D Ondusko, DF Hsu (2019). Improving portfolio performance using attribute selection and combination.
This paper proposes is a method to compute the attribute scoring system using weighted average by the recency giving more weight to scores at the time closer to the present. the researchers then show, by market testing, that their results perform better than that of Irukulapati et al. in a majority of the cases as well as the Russell 2000.

G. M Kis, S Gy (2017) Portfolio approach in higher education"
Portfolio has many interpretations rating and qualification of persons and also the training selection of an organization among others. In this paper, the researchers focus on portfolios relevant to the labor market. They examine whether they are suitable for collecting, organizing and presentation of learned knowledge, experience, competencies and relationships.

H. Srivannaboon, S. & Milosevic, D. Z. (2016). A Two-Way Influence between Business Strategy and Project Management, International Journal of Project Management, 24(6), 493-505.
This paper recognizes the strategic importance of project management in the corporate world through an exploration of PM/business strategy alignment. found that business strategy realizes its influence on PM via the competitive attributes of the business strategy time to market and quality cost and These competitive attributes are used deliberately to determine the configuration and emphasis placed on different PM elements Example are strategy and organization etc.

I. Walker, H. J., et al. (2014). Factors Influencing Organizational Change Efforts: An Integrative Investigation of Change Content, Context, Process and 75 Individual differences, Journal of Organizational Change Management, 20 (6) 761-773.
This study is one of the first to integrate factors common to all change efforts i.e. content and context and also process and individual differences. Further and it elaborates on how these factors interact to influence change success.

J. Moudhir Khalid Abdul Hameed, Nazar Khalaf Hussein (2018) Using Swarm Algorithm to Build Efficient Portfolio
This paper explores the decision centric view of the product development and implications for the Swarm Algorithm needed to fully punctuate the entire product life cycle. Inferences in this paper is based on the almost entirely on the results of research into how the people do design and make decisions.

K. Research portfolio analysis and topic prominence Richard Klavans, Kevin W. Boyack, Research portfolio analysis and topic prominence, Journal of Informetrics, Volume 11, Issue 4, 2017, Pages 1158-1174, ISSN 1751-1577.
In this paper the researchers address the issues of consistency and significance and also relevance when it comes to research questions and with the help of other researchers, intend to analyze a model of the 91,726 studies, or the study of forms), which includes more than 58 million documents. This paper presents a new indicator on the key issues here is an example of the profile of the speed, and, therefore, is required.

L. Data Analysis by Web Scraping using Python David Mathew Thomas, Sandeep Mathur, Amity Institute of Information Technology (2019)
This paper discusses how the web has become a serious problem as a result of the autonomous and heterogeneous nature of the hidden content of the world wide web. The traditional stress engine has now become an ineffective and inefficient way to search for such information. The most important outcome of this paper is to become a user-friendly interface for searching and indexing and also effective methods in order to obtain a data-based web architectures.

M. Electronic portfolios as a way to showcase technology skills (Heath, 2015)
The argument this paper makes is that electronic portfolios should be electronic versions of paper portfolios. The same thinking about purpose and pedagogy and assessment lies behind both kinds of portfolio general literacy ans also the communication and problem solving skills

©IJRASET: All Rights are Reserved

N. Web scraping technologies in an API world (2015) Daniel Glez-Peña, Anália Lourenço, Hugo López-Fernández, Miguel Reboiro-Jato, Florentino Fdez-Riverola

This paper discusses site-scraping, one of the most tried and true methods for the extraction of the content on the web, as well as its various benefits and the wide range of its applications. Scraping web data, which is defined by the contacts of the target website, using the HTTP text of the protocol is a stateless Internet protocol to coordinate the request of the efforts of the client, usually the web browser and the web servers. HTTP is typically the "method" used in order to RECEIVE requests a resource that is used in the polls, and according to your wishes, which is to be used when you submit the form, and upload the files. However, given that this has remained unchanged for many years, the paper examines whether this approach can be improved upon or not.

O. Mining Professional's Data from LinkedIn (2015) Fifth International Conference on Advances in Computing and Communications. Puneet Garg,Rinkle Rani,Sumit Miglani.

Social network is a web based service and it the allows users to create a public semi public or professional profile with in the domain such that they can connect with other individuals within the network.These days the analytics are focusing on observing the association and structure of various social networking sites. The main research area is to study and analyze compactness, centrality grouping with the promise of the security and privacy of the users data in these networks. LinkedIn is a predominantly appealing and exciting objective and specifying the proficient environment of its subscribers. By investigating LinkedIn networkinformation and data one can understand the behavior and other properties of individuals and groups. By collecting all the LinkedIn data that is occupation and employer and also designation etc. personality of the users can be exactly visualized. At first glance LinkedIn appears like other social websites due to the responsive and professional characteristics LinkedIn's data and its API are quite different from data available at other social network sites and have their own senses.

P. Data mining with LinkedIn Bradbury, D. (2020). Data mining with LinkedIn. Computer Fraud & Security, 2020(10), 5–8.

This paper discusses the two main ways to access LinkedIn via third party. The first is to use the LinkedIn Application Programming Interface while the second is to export your network data in a file and work on that. LinkedIn published gives API to give third party application developers access to its information. They can use the oauth authorisation protocol to let their own users sign in using their LinkedIn credentials. They can let their applications access those customers LinkedIn data. Accessing the LinkedIn API data is very easy to applying for a developer API token which gives to the ability to query the interface. You can then use the scripting language is your choice to access the data. One useful open source solution is LinkedIn project is provides Java wrapper for the LinkedIn API.

VI. METHODOLOGY

A. Using Artificial Intelligence for Recommendation

A recommendation system is generally an information filtering system that uploads information depending on the users interests, preferences, and previous history on an item. It can be able to recommend a specific user's choice on a product by their profile. With the use of AI Algorithm recommendation systems, the head hunters or interviewers are able to find the perfect candidate for the job easily and quickly. The recommendation engine is a splendid tool that we have incorporated in our project which will help highlight the job perfection that the user has for the role. This is especially useful because it will be personalised recommendations for each and every user based on his profile skills and experiences. The importance of recommendation engines in the resume builder is not only vital but also useful and innovative.

1) *Collecting the Data*: Data collection is the first step of creating a recommendation engine. In reality, data is Separated into explicit and implicit ones.data given by the users such as ratings and comments are clear. Whereas indirect data may include a search log, order and also return history, clicks, page views, and cart events. This type of data is gathered from any users who visit the given website.
2) *Storing the Data*: In order to have good recommendations, you need to create more data for the algorithms you use. That means you can change any recommended project to a good data project quickly . you may decide which type of storage you might want to use with the help of the data you use for creating recommendations. It is up to you whether to use a NoSQL database or maybe a standard SQL database or even some type of storage object.
3) *Analyzing the Data*: In order to look for items with similar user interaction data, it is important to change it with the use of various analyzing methods. Sometimes it is necessary to provide recommendations instantly when the end user is checking the item, so analysis is required.

©IJRASET: All Rights are Reserved

4) *Filtering the Data*: The next step is to filter the data to provide with suitable recommendations to the users. In order to take the use of this method, you should choose a algorithm that is most suitable for the engine you use. AI Algorithm used in our project for recommendation:

a) *Naive Bayes:* The Naïve Bayes algorithm is a classification technique based on applying Bayes' theorem with a strong feeling that all the predictors are not dependent on each other. In simple words, the assumption is that the presence of a feature in a class is usually independent even if there is any other feature present in the same class. Real time prediction- because its easy to implement and can compute fast it can be used to do prediction in real-time which will hence help us predict the best suited career for our user while he is on the website itself. Multi class prediction- this can be predicted by the Naive Bayes algorithm, posterior probability of multiple classes of target variable. Text classification − because of the feature of multi class prediction, Naive Bayes classification algorithms are well suited for text classification. So can also be used to solve problems in skill-filtering and role analysis. Recommendation system − Naïve Bayes makes a Recommendation system which can be used to filter unseen information and to predict whether a user would be eligible and suited for the given job or not. The steps are as follows:

- Input Training Dataset TD, where F = (f_1, f_2, f_3,.. f_n) are the vales of the predictor variable in the dataset
- Read the training dataset TD
- Calculate the mean and standard deviation of the predictor variable for each class
- Calculate P(f_i) using the Gauss Density Equation in each class
- i=i+1
- Repeat 4,5 until i=n
- Calculate the likelihood for each class
- Get the greatest likelihood

B. *Data Scraping for Collecting Data*

1) *Choose a Programming Language or Tool to Scrape Data*: You need to choose your tools for web scraping in a way that suits your needs. Python is the most popular as it has an easy learning curve, simple syntax, as well as the availability of multiple external libraries, and the language being open source. There are multiple libraries like Beautiful Soup and Scrapy that can be used to crawl web pages, create web crawlers, and run to scrape jobs at specific time gaps. Python allows for immensely flexible when it comes to combining other systems with our scrapping engine. We could be able to save our scrapped data easily on our local machine, in the databases, maybe on a storage or even may be dumped into a word or excel file.

2) *How to Scrape a Single Webpage and Identify the Components:* Before we get to scrape product information from a 1000 product pages for example on LinkedIn, you need to access the page and get the entire Html page to analyze it and decide on a strategy. Data in HTML pages can exist within specific key value pairs in tags or the text within tags. We can be able to use libraries like beautifulSoup to mention the exact tag you want to take the data from on all the webpages and run the code on a loop.

3) *Decide on a Data Cleansing and Storage Strategy:* Before you start to scrape the data, you should first decide where you will be storing it. This is because how you process the data will depend on where you will be storing it. There are multiple options available. You can choose between NoSQL and SQL databases, depending on weather the data that you scrape will be structured on unstructured. For data which is not in structure, you can choose SQL databases so that you can store the data in rows consisting of a set of attributes. For unstructured data, where the attributes are not set ,we may go to NoSQL database. Additionally, the dataset should be cleansed so that it can be used by the algorithm.

The data cleansing steps are as follows:

a) *Steps Involved in Scraping data from LinkedIn to our Website:* LinkedIn uses OAuth 2.0, an authentication and authorisation protocol, for authentication steps in order to configure our application.

The following is the standard procedure for scraping data from LinkedIn:

- Configure your LinkedIn Application (must be logged in already). Create a new application by mentioning all the required details and then submit.
- Once you submit, you will be directed to a page where you can configure your app. You will see a Client ID (API Key) and Client Secret (API Secret) values in the page. Note them down.
- Under default application permissions, you must select at least "r_basicprofile".

©IJRASET: All Rights are Reserved

- Next, under OAuth 2.0, you must specify a URL (redirect URL) that LinkedIn can call back. For this, you should be able to host a web page somewhere in the internet that is accessible to LinkedIn.
- Add url in the input box, click add, and then press the Update button at the bottom of the page.

b) The authentication mechanism itself works as listed below:
a) Use ClientID and Client Secret values, to obtain an authorization code.
b) Exchange the authorization code in order to get an access token.
c) Use the access token to retrieve data from your LinkedIn profile

C. Data Visualization:

This can be defined as a visual representation of a collection of related information in a meaningful manner, such as a chart or another image. The primary goal is to visualize data and statistics in such a way that the information is easily conveyed, as well as interpret the displayed data in order to infer or gain new information. Data visualization can be useful for data cleaning, as well as exploring data structure and detecting outliers and groups that are not usual, to identify trends and clusters, to spot local patterns, finding modeling output, and presenting results.

Visualizing the collected data from the user-

- Explains that the profile is suitable for a particular task or a job from the analyzed data we collected.
- Pie charts are ideal for illustrating proportions as we plan to show the best skill the user is well versed.
- It gives us a clear idea of what the information means by giving it visual context through maps or graphs.

For the purposes of our project, we've decided to use pie charts for visual Representation.

The main reasons are as follows:

- It displays relative proportions of multiple classes of data.
- The size of the circle can be made proportional to the total quantity it represents.
- Useful to summarize a large dataset in visual form.
- It is visually simpler than other types of graphs.
- Allows for a visual check of the reasonableness or accuracy of calculations.

Pie charts are mostly commonly used data visualization methods, used across a wide range of applications. This is built for visualizing how each part contributes to that whole and can only be used if the sum of the individual parts add up to a meaningful whole.

This makes the data more understanding to the human brain. So its easier to identify trends, patterns and outliners in large datasets.

D. Database

MongoDB is a widely used popular NoSQL database. It is a non-relational database, which means (unlike in the case of a SQL database) that it doesn't make use of tables for storing data, but rather uses JSON key-value pairs. While tables are a good way for structuring data and relating fields with one another, when data is distributed over multiple tables, it becomes tiresome attempting to perform analysis over the tables. In MongoDB, to overcome the data relationship over multiple collections, it has two ways for comparing different types of data — Embedding & Referencing. When we have data stored in some file formats we can import it into our database as collections and also export collections from a database to files. MongoDB supports JSON, CSV and TSV files for importing and JSON and CSV for exporting. We store the stuff like user's name, experience, the projects they worked on and the basic info about the user in our Database.

©IJRASET: All Rights are Reserved

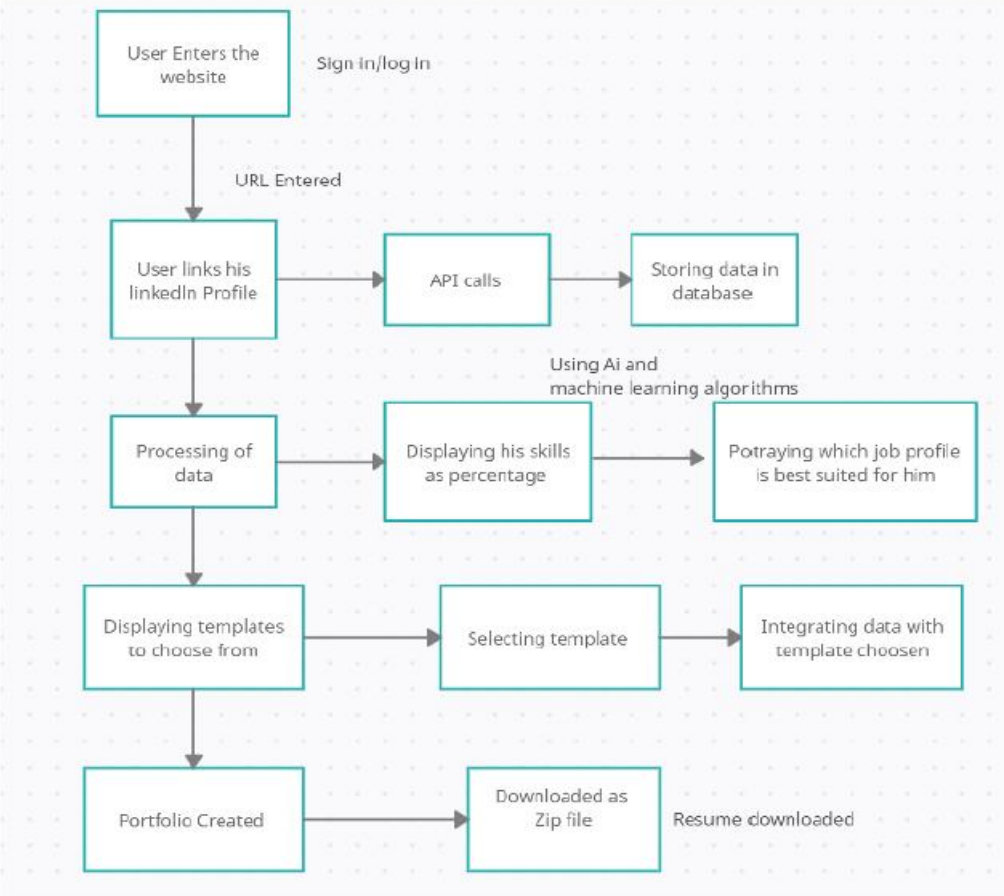


Fig. II Block diagram of our website

VII. FUTURE WORK

Our future aim is to add the following features to our website:

1) Custom drag and drop based UI.
2) Directly provide hosting for delivered websites.
3) Scraping of platforms other than LinkedIn like GitHub, GitLab etc to gather data.
4) Adding videos to the portfolio.
5) CLI interface for our website

VIII. CONCLUSIONS

We have built a website that will take data from a user's LinkedIn profile and build a custom portfolio using a variety of pre-developed designs. Provided method is an inexpensive and effective method to build highly attractive and customizable resumes. It also saves a lot of time by providing pre-designed templates. Our method also allows for non-technical background people to build cheap and high-quality web-based portfolios. Our application also reduces the use of paper in resumes by providing an eco-friendly web-based solution for building portfolios. It provides a platform to showcase the user's works with a custom-made portfolio with the user's choice of template. It makes users more accessible for the clients and companies and creates an online presence which further helps in reaching out to more people. We also provide a recommendation to the user based on their mentioned skills to which field of job he will be best suited for. We even give a visual portrayal of the user's skills through a pie chart which is easy to read and understand. Nowadays a portfolio is no longer a special thing it is required for every one as it shows who you are and what you are known for. Not all people can create their portfolio as there are a lot more fields apart from computers. Hence making creating a portfolio so easy and ready to use and even give the employers an insight on which domains the user can be best fit. This website will help users stand out from the crowd and get them one step closer to their dream job.

©IJRASET: All Rights are Reserved

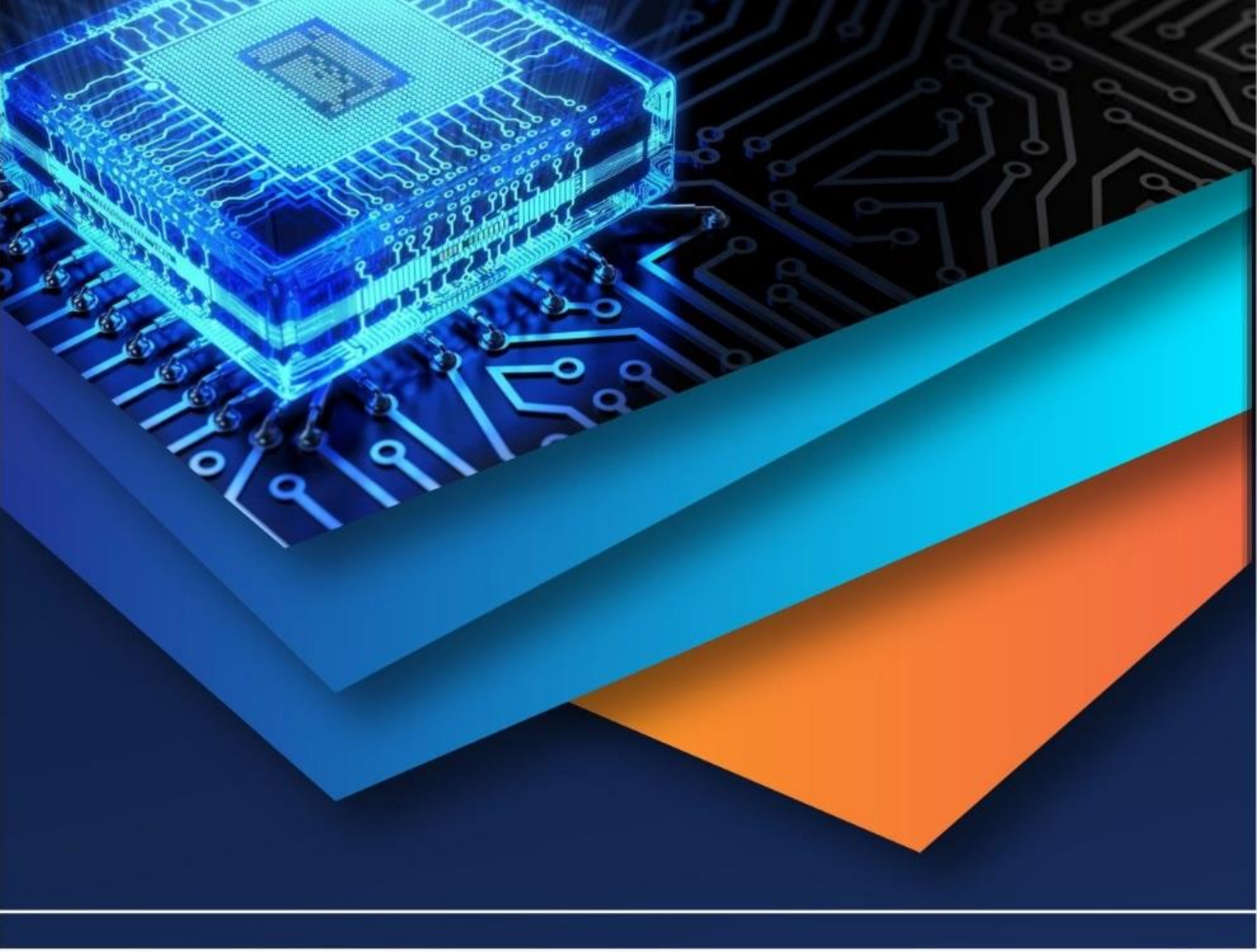
IX. ACKNOWLEDGMENT

We have taken a lot of effort into this paper. However, completing this paper would not have been possible without the support and guidance of a lot of individuals. We would like to extend our sincere thanks to all of them. We are highly indebted to Mr. Senthil Kumar K for his guidance, supervision, insightful leadership and knowledge which benefited us to complete this paper successfully. Thank you so much for your continuous support and presence whenever needed. We would like to express our gratitude towards our parents & our friends for their kind cooperation and encouragement which helped us a lot in completing this paper. Our thanks and appreciation also go to our colleague in developing the project. Thank you to all the people who have willingly helped us out with their abilities. We would also like to thank our college for providing us all the necessary resources for the paper. All in all, we would like to thank everyone involved in this project and helped us with their suggestions to make the paper better

REFERENCES

[1] Miles, M. and Huberman, A.M. (2010), Qualitative Data Analysis, 2nd ed., Sage, Thousand Oaks,CAJ. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
[2] Sullivan, S.E. (2009), “The changing nature of careers: a review and research agenda”, Journal ofManagement, Vol. 25 No. 3, pp. 457-84.M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, “High resolution fiber distributed measurements with coherent OFDR,” in Proc. ECOC’00, 2000, paper 11.3.4, p. 109.
[3] R. M. Amir Latif, M. Talha Abdullah, S. U. Aslam Shah, M. Farhan, F. Ijaz and A. Karim, "Data Scraping from Google Play Store and Visualization of its Content for Analytics," 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2019, pp. 1-8, doi: 10.1109/ICOMET.2019.8673523. (2002) The IEEE website. [Online]. Available: http://www.ieee.org/
[4] Fenwick, T. (2014), “Learning in portfolio work: anchored innovation and mobile identities”,Studies in Continuing Education, Vol. 26 No. 2, pp. 229-41.
[5] “PDCA12-70 data sheet,” Opto Speed SA, Mezzovico, Switzerland.
[6] J. Chen, H. Huang, S. Tian and Y. J. E. S. W. A. Qu, "Feature selection for text classification with Naïve Bayes", vol. 36, no. 3, pp. 5432-5435,
[7] M. Muniswamaiah, T. Agerwala and C. C. Tappert, "Federated Query processing for Big Data in Data Science," 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 6145-6147, doi: 10.1109/BigData47090.2019.9005530..
[8] Manoj Muniswamaiah, Tilak Agerwala, Charles C. Tappert, "Approximate Query Processing for Big Data in Heterogeneous Databases", Big Data (Big Data) 2020 IEEE International Conference on, pp. 5765-5767, 2020.
[9] M. Qamhieh, H. Sammaneh and M. N. Demaidi, "PCRS: Personalized Career-Path Recommender System for Engineering Students," in IEEE Access, vol. 8, pp. 214039-214049, 2020, doi: 10.1109/ACCESS.2020.3040338.Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1997.
[10] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," Knowledge-Based Systems, vol. 46, pp. 109–132, Jul. 2013.
[11] D. Richards and D. Terkanian, "Occupational employment projections to 2022," Monthly Labor Review, U.S. Bureau of Labor Statistics, Dec. 2013

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)