2020 Volume 1, Issue 1: 31 – 40 DOI: 10.48185/iitc.v1i1.94

Survival Analysis of Python and R within the Job Market Trend

Hina Afzal^{1,*}, Arisha Kamran², Asifa Noreen¹

¹Department of Computer Science and Technology, University of Lahore, Lahore, Pakistan ²Department of Computer Science and Technology, Capital University of Science & Technology, Islamabad (CUST), Pakistan

Received: 05.12.2020 • Accepted: 25.12.2020 • Published: 31.12.2020 • Final Version: 31.12.2020

Abstract: The market nowadays, due to the rapid changes happening in the technologies requires a high level of interaction between the educators and the fresher coming to going the market. The demand for IT-related jobs in the market is higher than all other fields, In this paper, we are going to discuss the survival analysis in the market of parallel two programming languages Python and R. Data sets are growing large and the traditional methods are not capable enough of handling the large data sets, therefore, we tried to use the latest data mining techniques through python and R programming language. It took several months of effort to gather such an amount of data and process it with the data mining techniques using python and R but the results showed that both languages have the same rate of growth over the past years.

Keywords: Python, R, job market, data mining, Survival analysis, social media.

1. Introduction

We live in a digital society and social medial play a vital role in the business world. All the companies put great interest in their presence on social media sites. It a sort of communication platform all around the world. Recruitment companies use it as a source of finding an efficient candidate. Surveys have found that while the candidate selection majority of hiring managers look over the candidate social media portfolio (1). So nowadays it is becoming essential for job seekers to be more actively part of for searching opportunity. The main aim of this paper is to detect the Job Market of Python and R. We gathered almost 30 papers related to python and R and collect several months of job ads from an online job agency. And compare these posts based on tags related to python and R jobs and compile the result for trending language in the market.

1.1. Python

Python is the versatile language used for back-end development, software development, data science, and developing system scripts among other things. It freely sources for mathematics and data analysis for statistics and analytics, therefore it became the language of premier for scientists and researchers. It is very easy to learn if you have a software engineering background. It is flexible in computing algorithms, scripting websites, making software, building apps, etc.

It supports multi-paradigm like object-oriented, structured, functional, aspect-oriented languages. In database handling, it can handle a large dataset. Python code can be written like a programming

^{*} Corresponding Author: ahina74@gmail.com

language. The main focus is to implement a bulk of instructions at once, it can also be executed by one line or clusters of instructions, making it unique and suitable for working with data collaboration. Packages and libraries of Python are Numpy, Pandas, Matplotlib, scipy, scikit-learn, tensor flow.

1.2. R

R appeared in 1992 as a programming language. It is a systematic language that works by splitting a programming task into a series of steps, plan, and rotations. This is a plus when it comes to building data models because it makes it relatively easy to understand how complex operations are carried out; however, it is often at the expense of performance and code readability(2). It emphasizes quality reporting concerning envision and frameworks for creating dynamic web applications R is used mostly by mathematical people and many people progress from spreadsheets to R. It is faster than python after 1000 iterations. The usability of R is vast it does not need of strong background of coding experience. It supports only procedural programming for some functions and object-oriented for other functions but it processes the problem of handling large data sets. R is easier at the beginning, but the complexity of advanced features makes it difficult. Packages and libraries of R are Ggplot2, tidyverse, caret.

The main focus of this paper is to analyze the job market trends related to these languages. Comparison of both languages the one language that rules them all for data science. Both are powerful tools each with their respective strengths and in general, it makes a lot more sense to think about how they compete in the market. For comparison purposes, we gather the bulk of data from different sites and locations. We consider different industries that hire people with Python, R, or both, includes finance, consulting, telecommunications, and software. Different tags are used for different categories like for R Data analyst, Data scientist, Manager, Scientist. Some of the job titles that require Python are Data architect, analyst, engineer, and Developer.

This paper serves the analysis of both languages concerning market trends by using the web mining approach. The remaining paper is mapped as follows:

The literature review is discussed in Section 2. Section 3 presents future aspects. Section 4 concludes the paper and outcomes of the analysis.

2. Literature Review

In this Portion, we review the literature on the analysis of python and R in the job market. Thoroughly we will discuss the literature analysis and web mining techniques to keep abreast of job market trends.

In [3], the comparative analysis of a single best tool does not identify for all circumstances. A dozen different open source packages are used to demonstrate the power of Python to increase the community of python developers at a higher rate. Python is a general-purpose language, it also encourages all types of the community for data analysis which increases the opportunity of packages without exploring it. In the testing, the highest time performances of R on the minimum dataset can be holding the data in RAM.

R is having a huge community with packages as explained in the analysis. Numerous logical and scholarly network communities keep on relocating to the utilization of R. The holding of data and process by using hard drive space is done by SAS. In the experiments, the Veteran programmer has written the SAS code and less experienced individuals have written the Python and R. Each researcher performing the coding is allowed to perform further validation of metrics. The

measurements would be relieved by the variety of involvement. A good Data Scientist ought not to be zeroing in on being a specialist in any single instrument The important decision should be made by many resources available for coding to become an expert.

In [4], we introduced our enhancements made in the Python QA framework, which plans to expand the information base and give better responses to the clients. The KB of the Python QA framework was improved with information from the network QA site Stack Exchange, which gives important Questions and Answers from the Python theme. The progressions made to the framework code permitted the consideration of a bigger information base. The Python QA framework has likewise benefited from the client's custom made data about the nature of answers present in the dataset. The data present in the appropriate responses, as a rule, has instances of Python code, turning the appropriate response progressively pertinent to the client. As future work, we can upgrade the Python QA framework by utilizing different CQAS, broadening the information base.

In [5], Stack Overflow as being one of the biggest Q&A people group the insufficiencies are confining engineers to ask abstract or open-finished inquiries, because of which the time isn't resolved to get a palatable reaction. The examination zeroed in on two language networks, R and Python, which are generally utilized by the associations as an innovation of decision, they will likely decide the term until an acknowledged answer was given in one or the other Python or R addresses be the utilization of Survival Analysis standards. The reaction seasons of the two dialects are comparative while contrasting the time until the acknowledged answer metric.

In [6], as expressed the capability of any examination is apportioned by the coming outcomes dependent on the emotional and complete investigation done on exploration information and its subtleties. Exploration references' in the wake of being encoded has gone magnificently to give remarkable outcomes to understand definitive targets of text mining as clarified. A progression of very much expressed differential results can be achieved in the way; Python and R programming have impregnated their methodology of taking care of crude unstructured information conquer the 3 significant year openings. This is a little reference that guarantees that how Python and R programming has contrasted top to bottom on the grounds of dissecting framework approaches for achieving various ends throughout some time. Eventually, the objective of understanding the content mining approach over some time utilizing differed philosophies has had the option to grow all around arranged results utilizing both Python and R programming, be that as it may, Python has built up an over- hand in taking care of circumstances and managing issues much because of its powerful executing speed, visual- explanations and unequivocally fathoming crude suits of information from plenty of exploration information accessible.

Table 1.

	Python Vs R Literature	review
year	Survival of python and R	Conclusion
2016	Business stick with java, python, and C	Comparison of different languages.

2018	COMPARISON N OF R VS. PYTHON FOR TEACHING COMPANY PROBLEMS	It related to the advantages of Python vs. R in teaching different types of students. And compare students from Math using R compared to using Python also contrast Python and R in solving problems for industries.
2017	DISCOVERING G JOB MARKET TRENDS WITH TE XT ANALYTICS	This paper examined work pattern investigation utilizing text examination. The objective was to gather, examine, and visualize valuable occupation market data that can be useful to a significant assorted scope of individuals. Mainly focus on the analysis of technical skills.
2018	USING TEXT MINING TO DISCOVER SKILLS DEMANDED IN SOFTWARE DEVELOPMENT NT	Finding expertise necessities of IT work market assist instructors with overseeing the educational program, plan courses, and preparing for creating understudy abilities which lead to improving the employability of graduates and fulfillment of employment.
2018	Comparison and contrast of Statistics Software Packages including R and Python for teaching purposes	Comparison of Python and R based on software compatibility.
2018	Extending Python QA with Knowledge from Stack Overflow	The Knowledge base of the Python QA framework was improved with information from the network QA site Stack Exchange, which gives applicable Questions and Answers from the Python theme.
2019	A Survey: How Python Pitches in IT-World	Python is timber to give a total arrangement in the realm of programming. It is an open-source language and can be utilized for some social and business objectives. It can likewise be utilized on any of the stages like working frameworks, for example, Windows, Mac, and Linux to give some examples.
2020	Machine Learning in Python: Main Development s and Technology Trends	It gave a concise foundation into significant points while researching The different difficulties and present status of answers for each.
2018	Survival Analysis within Stack Overflow: Python and R	It mainly focused on two language faction, most normally utilized by the associations as an innovation of decision, to attempt to decide the length until an acknowledged
		answer was given (assuming any) in one or the other Python or R addresses utilizing Survival Analysis standards.

2018	Python Developer Survey 2018	In 2018 JetBrain has conducted an official survey on Python Developers for finding the strength of Python developers in the market. More than twenty thousand developers were to map out the Python community among more than 150 countries.
2019	Python for Data Analytics, Scientific and Technical Applications	Although Python is slower in runtime and has some plan limitations when contrasted with aggregated dialects like C or C++, it is favored by researchers and engineers in the field of information investigation, mathematical calculations, and practically all specialized areas like Machine learning, AI, Deep learning, and so on.
2018	Packaging and Distributing Your Python Project to PyPI for Installation Using pip	This tutorial will explain the means needed to bundle your Python ventures, disseminate them in circulation designs utilizing step devices, transfer them into the Python Package Index (PyPI) repository utilizing twine, lastly establishment utilizing Python installers, for example, pip and conda.
2017	Best practices of orchestrating Python and R code in ML projects	Python and R both languages vastly used in data science projects. Extra apparatuses other than git and shell scripting are created to encourage the improvement of the prescient model in diverse-language conditions.
2016	Data Science Programming g: Python vs R	The best arrangement is to take a savvy action dependent on the above- recorded techniques and choose which language you ought to learn first that will bring you employment with a major information researcher pay and later add to your range of abilities by learning the other language
2016	R vs Python: head to head data analysis	We've investigated how to dissect a dataset with R and Python. There are numerous undertakings we didn't jump into, for example, continuing the aftereffects of our examination, offering the outcomes to other people, testing and preparing things creation, and making more representations.
2020	Compiler Design Life Cycle and Comparison of Programming Languages	Python language comparison with the language of C# we found there is a difference where the Python language was distinguished by the ease of the code and the size of the code is very small.
2018	Spatial data analysis with R programming g for environment	In open foundations or regions, development plans, metropolitan data frameworks, and other data frameworks can be made and it permits the occasion to be introduced on the web to the clients complimentary.
2018	The Usage of R Programming g in Finance and	R programming language is comparatively new from Python in the domain of data analytics. It is new for data

	Banking Research	analytics despite this it has several packages for solving the fiancé problems. These packages will continue to still
2016	An Interactive Web-tool for ROC Curve Analysis Using R Language Environment	Since ROC bend examination is one of the central measurable investigation techniques, it is utilized by a wide scope of established researchers. Both business and free programming devices are accessible for clients to perform it. For the most part, simple to-utilize and pleasantly planned interfaces are offered by business programming bundles while adaptable and extensive instruments are accessible in free, open access, codebased programming bundles, for example, R.
2020	Development of "Physical Parametrization with Python"	It has been effectively used to recognize the wellsprings of the time-step reliance which were available in the microphysical plot being used in the AROME and Meso-NH models.

In [7], in this paper, we have analyzed the potential gains and drawbacks of Python, MatLab, and R. We have investigated all of the lingos to one another, while moreover examining the informative assessment of each program in an instructing circumstance. In the wake of investigating all of the three of the tasks all around, we have shown up at the goal that Python is the best language to be told in an examination lobby condition. This is because it is not hard to use and will allow understudies admittance to open-source coding that can be found online when playing out a dynamically inconvenient assessment. Regardless, we should observe that R might be a better program than train to understudies since it is extensively used in associations around the nation for data mining. Thinking about a program like R could outfit understudies with a high ground while looking for a business upon graduation. Despite R, SPSS or SAS could in like manner be viable decisions that should be seen when indicating huge data assessment to understudies.

In [8] this paper Authors explain, the advances conducted in SW by misusing ML techniques overviewed. Explicitly image-based and numeric-based arrangements that broken down featuring their principal characteristics and disadvantages. Thus the fundamental imagined research bearings had drawn.

In [9], in this paper, we have proposed an astute framework to prescribe suitable positions to specialists from various online commercial centers. A potential occupation list is made for the specialists by coordinating these continuous ranges of abilities with the aptitudes needed in the posted positions. Our proposed keen framework won't just assistance the consultants secure their appropriate positions effectively yet will likewise help the requesters in assigning out positions to reasonable up-and-comers.

In [10], this paper presents an effective web personalization approach contingent upon the intermittent openness and web utilization log mining. It will bolster occasional web personalization using singular clients' intermittent perusing designs. Not at all like the occasional methodology, the proposed approach permits clients to capably settled which assets they are predominantly worried about for a given period selective of using the client's current availability data. Subsequently, it can perform customized asset provisioning that is all the more exorbitant in ridiculous situations. The testing assessment results recommended that the proposed approach very viable in wording client

appropriateness and palatable measure in different time stretch using web use log. In the future, it very well may be assessed in a more drawn out span in various periods to gauge the adequacy.

In [11], the proposed structure for Job Recommendation System is pointed toward overcoming any barrier among workers and bosses. Moreover, further prospects are as follows: Implementing Partner Program of LinkedIn, which requires banding together with LinkedIn for future tasks. Utilizing this service every API can be accessed by the developers for users. This will help make ability and supports planning simpler and confirmed.

In [12], these Authors, present two honesty check methods for redistributed incessant item set mining. The probabilistic confirmation method builds proof of frequent itemsets. Specifically, it expels little arrangement of things from the first dataset and addition a little arrangement of fake exchanges into the data to build proof persistent itemsets. The inevitable methodology needs a different build for cryptographic verifications of the mining result. The rightness and fulfillment are estimated in the case of the evidence with 100% conviction. Our examinations convey the proficiency and viability of our methodologies. A fascinating course to investigate is to stretch out the model to permit the customer to indicate her confirmation needs regarding financial plan (potentially in the fiscal arrangement) other than accuracy and review edge.

In [13], in this paper, we explored endurance examination, which is otherwise called occasion history investigation. For multivariate information, an endurance examination can turn out to be rapidly mind boggling, and the translation of it nonpaltry, particularly as for peril proportions. Consequently, we focused on a complete yet mixed introduction that is sufficiently wide to contain all points required for a multivariate endurance investigation yet is simultaneously comprehendible. Supplemented by a worked model utilizing R.

Table 2.

Year	Web mining approaches	Techniques
2020	Machine Learning for the Semantic Web: Lessons learned and next research directions	In this paper, the most delegate recommendations, having a place with the previously mentioned classifications are studied, together with the investigation of their fundamental quirks and disadvantages.
2019	Development of an Intelligent Job Recommender System for Freelancers using Client's Feedback Classification and Association Rule Mining Techniques	In this paper data mining approach classification and Association, rule mining is used to find out the skillsets of freelancers for achieving their job goals.
2017	An Efficient Web Personalization Approach based on Periodic Accessibility and Web Usage Mining	A well-organized web personalization method based on user browsing time interval and utilizing web usage log is proposed.
2019	Job Recommendation A system based on Machine Learning and Data Mining Techniques using RESTful API and Android IDE	For the "Job Recommendation System - Vitae" used ML and DM approach with REST ful web server application that makes the connection among server and client.

3. Future Aspects

Data analysis as part of this study form on different social media sites. Base information includes the different aspects includes tags, posts, views, post history, and links. This study mainly focuses on two programming languages python and R. Both are compeer languages in the market.

After the comparison of both languages, it may open new ways for researchers. In the future, they can predict the latest trend of programming language by using advanced web mining techniques. Also, Software development companies are increasingly great courage for online social media sites for posting ads. Students can easily pick their future directions in the IT industry.

4. Conclusion

This study intimated that data mining methods could be used by organizations and job seekers for the analysis of marketing trends. And will be able to furnish the result of this analysis to the platforms for future action. Also, to illustrate the execution and analysis is performed for two programming languages python and R. the analysis shown that both are very powerful languages. Both languages have shown remarkable growth in the last few years. We found that the most broaden of python can be explained by the amplification of data science and R is primarily used in statistical analysis. Although it's up to the individual data scientist or data analyst to pick the language that best accommodates their special needs. The Following are the reasoning of the conducted study:

- Python and R are both extraordinary programming languages for data science and analytics. Since they're open-source, they're free for everybody, not at all like commercial tools like SAS and SPSS. Either language is reasonable for practically any data science task, from information control and robotization to impromptu analysis and exploring datasets. Many may use the two programming languages for various purposes, e.g., conducting early-stage data analysis and exploration in R, then switching to Python when it's time to ship some data products.
- Python has a strong ecosystem and is regularly viewed as one of the simpler programming languages to peruse and learn. Its programming linguistic structure is straightforward and its orders mirror the English language, E.g. print ("Hello world!"). Ultimately, Python code is linguistically clear and exquisite, effectively interpretable, and simple to type. It's extraordinary for building data science pipelines and AI procedures, which can be incorporated with web structures at scale. On Contrast, R has a rich ecosystem of cutting-edge interface packages available to communicate between open-source languages. This permits the programmer to string their work processes together, which is particularly valuable for data analysis.
- Presently since the IT business sincerely values adaptability, Python is adaptable for making something that has never been done before. Programmers can likewise utilize it for scripting websites or different applications. Then again, R is just used to orchestrate statistical models.
- No company can profit from a programming language if there is no developer ready to learn it due to its intricacy, so a programming language ought to be anything but difficult to learn. Python's

attention to coherence and straightforwardness implies its expectation to absorb information that is moderately direct and smooth. Python is viewed as a decent language for novice software engineers. Though R is simpler to realize when you begin, yet the complexities of cutting-edge functionalities make it progressively hard to create aptitude. R isn't difficult for experienced software engineers to learn.

• Since Data visualization is a big thing nowadays Python has great packages for data visualization: seaborn: Library based on Matplotlib, Bokeh: for visualization library

Pygal: Create dynamic SVG charts

For individuals with no software engineering experience, R can be progressively hard to learn because it was created by analysts, not to make coding simpler. In any case, R has a lot of packages, which gives ground- breaking yet simple-to-learn tools for importing, manipulating, visualizing, and reporting on data. Finding the correct packages to use in R might be tedious and there are different limitations in the libraries of R.

- Python is for the most part utilized when the information examination errands should be coordinated with web applications or if insights code should be fused into a creation information base. Since it's an irrefutable programming language, Python is a fair instrument to execute calculations for use underway. R is mostly utilized when the information examination undertakings require independent figuring or investigation on individual workers. For exploratory work, R is simpler for tenderfoots. Measurable models can be composed of a couple of lines of code.
- Python users have to install the packages according to their analysis requirements. There are varieties of packages in python. NumPy, Pandas, Matplotlib, sci- kit-learn are the most popular libraries of Python. R packages having several functions and data sets. for example for large datasets probably use the dplyr and data.table.
- There are many Python IDEs to choose from which sharply maintaining code, Results, and storage of files. Jupyter Notebooks and Spyder are popular, and Jupyter Lab is gaining traction. Tip: Also try Rodeo, the data science IDE for Python.RStudio is the most popular R IDE. It's available in two formats: RStudio Desktop for running on the desktop application and RStudio Server for access via a web browser.

References

- [1] NordicJobsWorldwide2019:"how to use soial media in job market"https://www.nordicjobsworldwide.com/how-to-use-socialmedia-on-the-job-market/
- [2] Data-Driven Science (DDS) provides training for people building a career in Artificial Intelligence https://medium.com/@data_driven/python-vs-r-for-data-science- andthe-winner-is-3ebb1a968197
- [3] Brittain, Jim; Cendon, Mariana; Nizzi, Jennifer; and Pleis, John (2018) "Data Scientist's Analysis Toolbox: Comparison of Python, R, and SAS Performance," SMU Data Science Review: Vol. 1: No. 2, Article 7.
- [4] Azevedo, Renato Preigschadt de, Henriques, Pedro Rangel, Pereira, Maria João (2018) "Extending pythonQA with knowledge from stackOverflow".
- [5] Laurel Lord, John Sell, Feyzi Bagirov, Mark Newman (2018) "Survival Analysis within Stack Overflow: Python and R"
- [6] Neeraj Bhanot, Harwinder Singh, Divyansu Sharma, Harshit Jain, Shreyansh Jain less (2019) "Python vs. R: A Text Mining Approach for analyzing the Research Trends in Scopus Database"
- [7] Ceyhun Ozgur, Taylor Colliau, Grace Rogers, Zachariah Hughes, Elyse "Bennie" Myer-Tyson (2017) "MatLab vs. Python vs. R"
- [8] Claudia d'Amato, Pascal Hitzler, Manhattan, Krzysztof Janowicz, Santa Barbara (2020) "Machine Learning for the Semantic Web: Lessons learnt and next research directions"
- [9] Md. Sabir Hossain, Mohammad Shamsul Arefin (2019) "Development of an Intelligent Job Recommender System for Freelancers using Client's Feedback Classification and Association Rule Mining Techniques"
- [10] Y.Raju, Dr. D. Suresh Babu and Dr. K. Anuradha (2017) "An Efficient Web Personalization Approach based on Periodic Accessibility and Web Usage Mining"

- [11] Harsh Jain, Misha Kakkar (2019) "Job Recommendation System based on Machine Learning and Data Mining Techniques using RESTful API and Android IDE"
- [12] Suraj Shinde, Akshay Mohite, Sandip Shinde, Prof.Mangesh Manke (2017) "Verifying and Correctness of Frequent Itemset Minining"
- [13] Frank Emmert-Streib and Matthias Dehmer (2019) "Introduction to Survival Analysis in Practice"