



# SEARCH JOB ROLES WITH RIGHT SET OF SKILLS USING DATA ANALYSIS AND VISUALIZATION SYSTEM-SKILLSYNC.

<sup>1</sup>Asma Taj H A, <sup>2</sup>Shekh Md moinuddin, <sup>3</sup>Syed shariq kamran, <sup>4</sup>Selim jhangir, <sup>5</sup>Murari kumar

<sup>1</sup>Assistant Professor, <sup>2</sup>Student, <sup>3</sup>Student, <sup>4</sup>Student, <sup>5</sup>Student,

<sup>1</sup>Information Science and Engineering,

<sup>1</sup>Cambridge Institute of Technology, Bengaluru, India

**Abstract:** SkillSync is the bridge that connects talent to opportunity, offering an open-source platform where skills are showcased, discovered, and perfectly matched with the ideal job roles. This project seeks to revolutionize the way we approach the workforce, providing a plethora of benefits, including enhanced efficiency, reduced costs, and an expansive network of skills that now have the chance to shine

**Index Terms** - Component, formatting, style, styling, insert.

## I. INTRODUCTION

In a world where the professional landscape is ever-evolving, the ability to efficiently match the right skills with the perfect job role is a constant challenge. SkillSync is a groundbreaking project that has emerged as a beacon of innovation in the realm of talent discovery and skill matching. By combining cutting-edge technology with a commitment to simplifying and streamlining the traditionally complex processes of talent recruitment and job seeking, SkillSync offers a dynamic solution for job seekers, employers, and HR professionals alike.

In the following exploration of SkillSync, we will delve into its core components, dissect the challenges it addresses, and uncover the myriad advantages it brings to the forefront of the professional world. SkillSync is more than a project; it's a vision for a future where talent finds its true calling, and opportunities find their ideal Core Objectives of SkillSync:

For this study secondary data has been collected. From the website of KSE the monthly stock prices for the sample firms are obtained from Jan 2010 to Dec 2014. And from the website of SBP the data for the macroeconomic variables are collected for the period of five years. The time series monthly data is collected on stock prices for sample firms and relative macroeconomic variables for the period of 5 years. The data collection period is ranging from January 2010 to Dec 2014. Monthly prices of KSE -100 Index is taken from yahoo finance.

## II. LITERATURE SURVEY

1. Richard Fedorko, Stefan Krail, and Radovan Bacik, "Artificial Intelligence in E-commerce: A Literature Review". With the development of the information and communication technologies, artificial intelligence is becoming increasingly popular. The main aim of companies in today's e-commerce world is to influence customer behaviour in favour of certain products and brands. The application of artificial intelligence as an innovative tool in the field of e-commerce may seem as a positive step forward.
2. Jeevan S Devagiria, Sidike Pahe inga, Quamar Niyazb, Xiaoli Yangc, Samantha Smith, "Augmented Reality and Artificial Intelligence in Industry: Trends, Tools, and Future Challenges". Augmented Reality (AR) is an augmented depiction of reality formed by overlaying digital information on an image of objects being seen through a device. Artificial Intelligence (AI) techniques have experienced unprecedented growth and are being applied in various industries. The combination of AR and AI is the next prominent direction in upcoming years with many industries and academia recognizing the importance of their adoption.
3. Apoorva Jain, Assistant Professor, Department of Computer Science Engineering, JEMTEC, Greater Noida, "Study of Artificial Intelligence (AI) with Augmented Reality (AR)." Artificial Intelligence is the subarea of Computer Science, which deals with making machines work like humans. The concept of artificial intelligence has almost overruled the concept of typing on various Search Engine tools, rather people are interacting with computers like they interact with the humans. They give command to the systems and the systems respond accordingly, as humans does.

## III. METHODOLOGIES

When searching for job roles that require skills in data analysis and visualization systems methodology, you'll want to look for positions that involve the following key aspects:

- Data Analyst
- Business Intelligence Analyst
- Data Scientist
- Data Engineer
- Visualization Specialist
- Business Analyst
- Market Research Analyst
- Financial Analyst
- Operations Analyst

When searching for job listings, look for descriptions that mention requirements or responsibilities related to data analysis, visualization, statistical analysis, and relevant software tools. Additionally, consider the industry and specific domain expertise required for the role.

### Naive Bayes Algorithm

Naive Bayes algorithm is commonly used in data analytics for classification tasks. Here's how it is typically applied:

**Classification:** One of the primary uses of Naive Bayes in data analytics is for classification tasks. It's particularly effective in scenarios where you have a large number of features (attributes) and a relatively small dataset.

**Text Classification:** Naive Bayes is widely used in text classification tasks such as spam detection, sentiment analysis, and document categorization. In these tasks, the algorithm calculates the probability

that a given document belongs to a particular category based on the occurrence of words or features in the document.

**Feature Selection:** Naive Bayes can also be used for feature selection. Since the algorithm assumes independence between features, it can help identify the most relevant features for a given classification task.

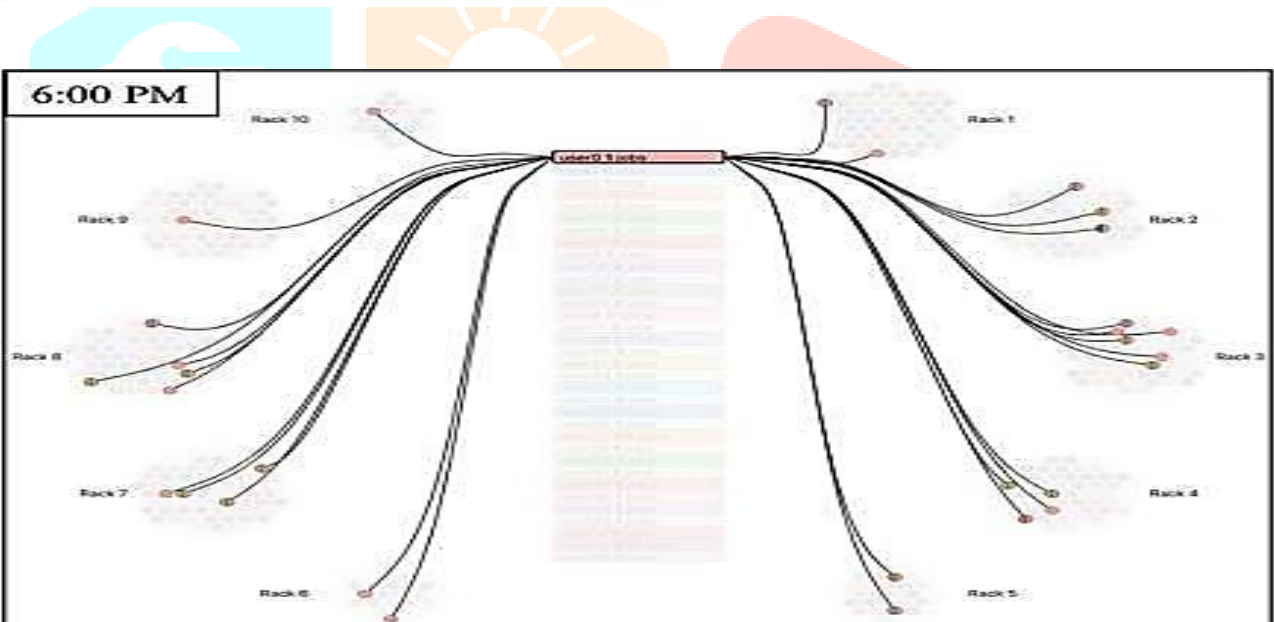
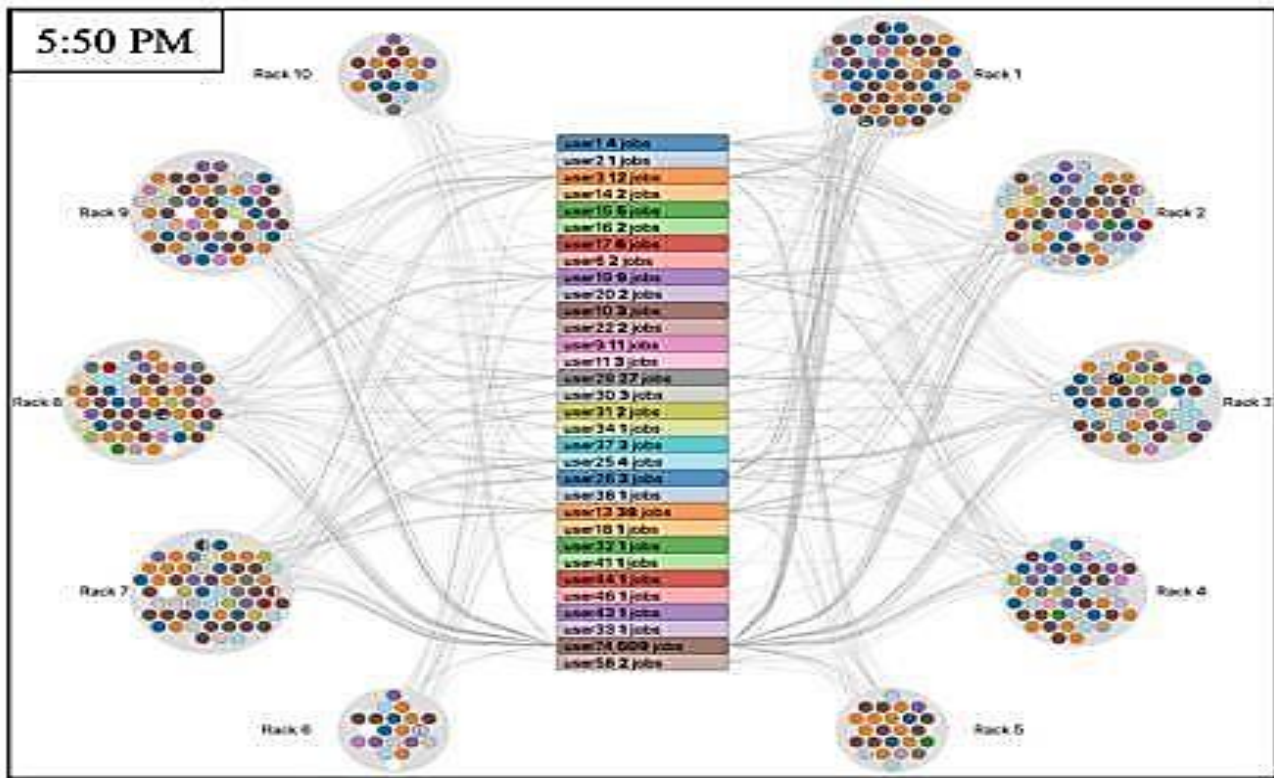
**Real-time Prediction:** Naive Bayes is computationally efficient and can make predictions quickly. This makes it suitable for real-time prediction tasks such as recommendation systems and customer segmentation.

**Imbalanced Data:** Naive Bayes can handle imbalanced datasets well, where one class is significantly more frequent than the others. It achieves this by using class priors to adjust for the class imbalance. It can be of two types:

1. **Categorical Variable Decision Tree:** Decision Tree which has a categorical target variable then it called a Categorical variable decision tree.
2. **Continuous Variable Decision Tree:** Decision Tree has a continuous target variable then it is called Continuous Variable Decision Tree.

#### IV. JOB ALLOCATION

The first use case focuses on how Job Viewer provides information about job allocations. Figure 2.a shows a snapshot of the main visualization on 08/14/2020 at 5:50 PM. The color distinguishes between different users, along with their related compute nodes. If a compute node runs several jobs of multiple users, it has all corresponding colors. If a compute node is white, no user's job is running on it. At 5:50 PM, there are nine white compute nodes that locate in six different racks. Ten minutes later, user0's job starts, as highlighted by the black outline and links in figure 2.b. It takes 1080 cores, or 30 compute nodes (each compute node has 36 cores). The system allocates seven out of nine white compute nodes to this job, and there are still two white compute nodes at 6:00 PM. One is on rack 2, and another is on rack 9. Figure 2.c highlights all 30 compute nodes running the user0's job. From these 30 compute nodes, 18 ones run two jobs, and 12 others run only one job. We have checked and found that most of the 18 compute nodes' former jobs consume all 36 cores at 5:55 PM. It means some of the compute nodes utilize up to 72 cores, including virtual cores, at 6:00 PM. These figures show information about job scheduling. Although there are two unused compute nodes, and the job requires so many cores to run, the system reuses the compute nodes running another job and does not allocate the two unused ones to the job. This use case is an example that can illustrate how efficient Job Viewer can help HPC administrators to monitor job scheduling.



### 3.1 RELATION BETWEEN JOB AND HEALTH STATE

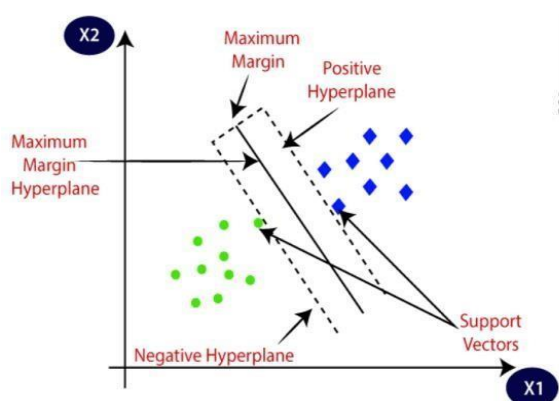
This use case clarifies the relations between jobs and the health states of compute nodes. We firstly look at the time series of CPU2 temperature of the compute node 4.33 in figure 5. The unit of temperature is degree Celsius, and the time takes place in August 2020. The vertical dash line indicates the time step at which we stop the timeline to get the time series. It is 08/18/2020 at 11:30 AM when we investigate the previous use case. The colorful areas highlight periods when a job is running on the compute node. We use text notations, which have similar colors to the corresponding Figure 4: The visualization of all compute nodes in rack 4. The color indicates (a) CPU1 temperature and (b) CPU2 temperature. Red color means high value, yellow depicts a medium temperature, and green represents low value. Figure 5: The time series of CPU2 temperature of compute node 4.33. The colorful areas highlight the period when the user with similar color runs his/her job.



### 3.2 Naïve Bayes Classifier

The naive bayes classifier calculates the probability of a file belonging to a specific category depending on several metrics obtained from the training dataset. A file with a high probability to be clean is begin ware, else if the probability is low.

- **Formatting data into a Pandas readable format**



- **Gaussian Radial Basis Kernel Support Vector Machine**

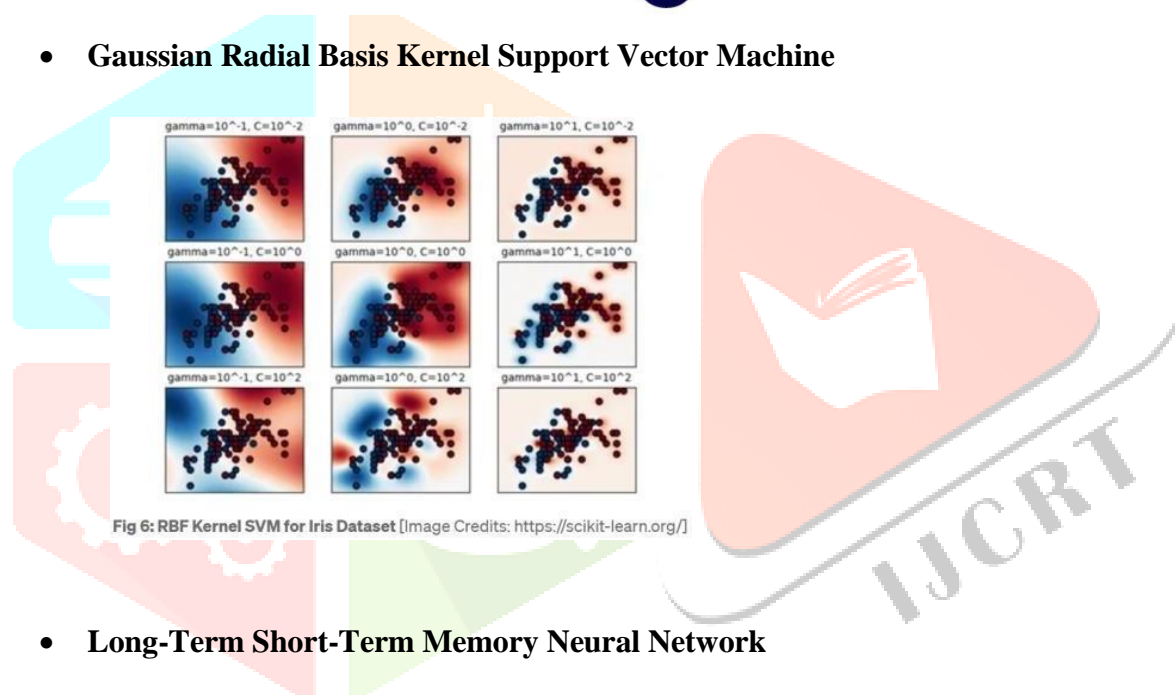


Fig 6: RBF Kernel SVM for Iris Dataset [Image Credits: <https://scikit-learn.org/>]

- **Long-Term Short-Term Memory Neural Network**

$$K(X_1, X_2) = \exp\left(-\frac{\|X_1 - X_2\|^2}{100}\right)$$

### Detection Time

To highlight the performance of ML detectors, the same binaries were tested using the different classifiers. Figure 7 summarizes the detection time of each classifier. For a same new binary to test, the neural network and logistic regression classifier achieved the fastest detection rate (4.6 seconds) and the random forest classifier the slowest average (16.5 seconds).

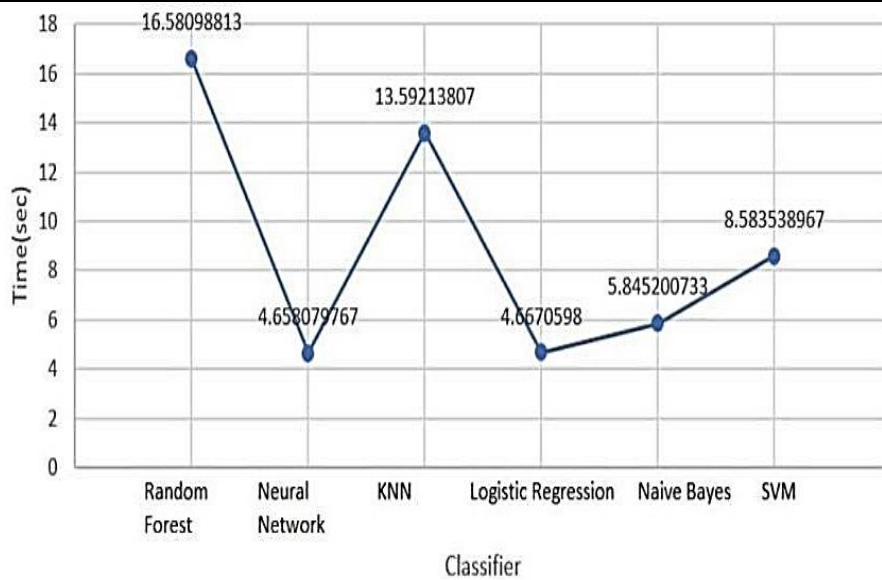


Fig 6: Average Detection Time

### 3.3 CONCLUSION

In conclusion, job roles that require skills in data analysis and visualization systems methodology are crucial in various industries for extracting insights from data and making informed decisions. These roles typically involve proficiency in tools such as Python, R, SQL, Tableau, Power BI, or similar, along with a strong understanding of statistical analysis and data visualization principles.

Effective communication skills are essential for conveying insights to stakeholders, and domain knowledge may be necessary depending on the industry. Job titles such as Data Analyst, Business Intelligence Analyst, Data Scientist, and Visualization Specialist are common in this field.

### 3.4 REFERENCES

- [1] W. Allcock, E. Felix, M. Lowe, R. Rheinheimer, and J. Fullop. Challenges of hpc monitoring. In SC'11: Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 1–6. IEEE, 2011.
- [2] N. Barlow and L. J. Stuart. Animator: A tool for the animation of parallel coordinates. In Proceedings. Eighth International Conference on Information Visualization, 2004. IV 2004., pp. 725–730. IEEE, 2004.
- [3] S. Benedict. Performance issues and performance analysis tools for hpc cloud applications: a survey. *Computing*, 95(2):89–108, 2013.
- [4] E. Betke and J. Kunkel. Real-time I/o monitoring of hpc applications with siox, elastic search, Grafana and fuse. In International Conference on High Performance Computing, pp. 174–186. Springer, 2017.
- [5] D. Carasso. Exploring splunk. CITO Research New York, USA, 2012.
- [6] T. N. Dang, A. Anand, and L. Wilkinson. Timeseer: Scagnostics for high-dimensional time series. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):470–483, 2012.
- [7] A. Dasgupta, R. Kosara, and L. Gosink. Meta parallel coordinates for visualizing features in large, high-dimensional, time-varying data. In IEEE Symposium on Large Data Analysis and Visualization (LDAV), pp. 85–89. IEEE, 2012.
- [8] N. Enterprises. Nagios. website.

- [9] Grafana. The open platform for beautiful analytics and monitoring, 2019. <https://grafana.com/>.
- [10] J. A. Hartigan. Clustering Algorithms. John Wiley & Sons, Inc., New York, NY, USA, 99th ed., 1975
- [11] H. R. Hartson. Human–computer interaction: Interdisciplinary roots and trends. Journal of systems and software, 43(2):103– 118, 1998.
- [12] N. D. Hugh Greenberg. Tivan: A scalable data collection and analytics cluster. 2018. The 2nd Industry/University Joint International Workshop on Data Center Automation, Analytics, and Control (DAAC).
- [13] A. Inc. Amazon cloudwatch, 2012. <http://aws.amazon.com/cloudwatch/>..
- [14] M. L. Massie, B. N. Chun, and D. E. Culler. The ganglia distributed monitoring system: design, implementation, and experience. Parallel Computing, 30(7):817–840, 2004
- [15] G. Palmas, M. Bachynskyi, A. Oulasvirta, H. P. Seidel, and T. Weinkauff. An edge-bundling layout for interactive parallel coordinates. In 2014 IEEE Pacific Visualization Symposium, pp.

