

# A Data-Driven Study on Emerging Skill Demand in Online Job Portals using NLP and Time-Series Forecasting

Rishav Kumar

## Abstract

This study explores the dynamic nature of skill demand in the online job market by leveraging real-time job posting data. A dataset was constructed through web scraping of job listings from Internshala, capturing key attributes such as job descriptions, required skills, and posting dates. Natural Language Processing (NLP) techniques were applied to extract relevant skill information from unstructured text data.

The extracted data was analyzed to identify temporal trends in skill demand, followed by the application of time-series forecasting models to predict future demand. A comparative evaluation between keyword-based extraction and TF-IDF methods was conducted to determine effectiveness. The findings indicate a significant rise in demand for skills such as Python, data analytics, and cloud technologies. The proposed approach demonstrates the potential of combining NLP and predictive analytics for understanding and anticipating job market trends.

## Keywords

Job Market Analysis, NLP, Skill Extraction, Time-Series Forecasting, Web Scraping, Data Analytics

## Introduction

The modern job market is highly dynamic, especially in the field of technology, where skill requirements evolve rapidly. Students and job seekers often struggle to identify which skills are in demand and how those demands change over time. Traditional sources such as surveys and reports are often outdated and fail to capture real-time trends.

With the rise of online job portals, a large amount of real-time job data is available. This data provides an opportunity to analyze current trends and predict future skill requirements. However, job descriptions are typically unstructured, making it challenging to extract meaningful insights.

This paper proposes a data-driven framework that leverages web scraping, Natural Language Processing (NLP), and time-series analysis to identify and predict emerging skill demands.

## Literature Review

Previous research has used web scraping to collect job postings and analyze demand for specific roles. NLP techniques such as TF-IDF have been applied to extract relevant information from job descriptions. Time-series models like ARIMA have been used to forecast trends in various domains.

However, most studies focus either on skill extraction or trend analysis separately. This paper integrates both approaches and adds predictive analysis for future skill demand.

## Methodology

Data Collection: Job postings were scraped from Internshala including job title, description, skills, location, and date.

Data Preprocessing: Cleaning, tokenization, stopword removal, and normalization.

Skill Extraction: Keyword-based matching and TF-IDF methods were used.

Trend Analysis: Skill frequency tracked over time.

Prediction: Moving average and ARIMA models applied.

Visualization: Graphs created for insights and forecasting.

## Results and Discussion

Top skills identified include Python, SQL, Data Analysis, and Machine Learning. Increasing trends were observed for Python and AI-related skills. ARIMA model showed better prediction performance than moving average. TF-IDF outperformed keyword extraction in accuracy.

These findings highlight the effectiveness of data-driven analysis in understanding job market dynamics.

## Conclusion

This paper presented a framework for analyzing and predicting job market skill demand using NLP and time-series techniques. The results demonstrate that real-time job data can provide meaningful insights into evolving skill requirements. The approach can assist students and professionals in making informed decisions.

## Future Work

Future improvements include using advanced models like BERT, expanding datasets, and building real-time dashboards.

## References

1. Salton, G., & Buckley, C. (1988).
2. Box, G. E. P., & Jenkins, G. M. (1976).
3. Bird, S., Klein, E., & Loper, E. (2009).