Thanapoom Phatthanaphan

Data Analyst Proficiency Insight: A Comprehensive Skills and Qualifications Analysis

1. Introduction

In today's job market, there's a growing need for people who are good at data analysis. This demand is especially high in sectors like Banking, Media & Entertainment, and Healthcare, where working with big sets of data is common. Many individuals are keen on starting a career in technology, specifically as a data analyst.
However, becoming a data analyst involves learning various skills. To make things a bit tricky, different industries look for different skills. For those who are just starting out, figuring out which skills are most important can be a bit confusing.
Our project is here to help with that. We're collecting information from popular job websites and carefully studying job descriptions to figure out the main skills you need to be a successful data analyst. The goal is to make things easier for people who are just starting out, giving them a clear idea of the skills that really matter in the real world of data analysis jobs. This project is all about helping newcomers make smart choices when it comes to developing the skills they need to succeed in the field.

2. Preliminary Literature Review

Numerous studies have outlined methodologies for analyzing job descriptions to pinpoint the requisite skills for specific positions. For instance, the research named "Skill Requirements Analysis for Data Analysts Based on Named Entities Recognition," as published by Lina, C. and Jian, Z. (2021), conducted an in-depth exploration into skill requirements for data analysts, utilizing job advertisements from 51job.com. Employing crawlers and manual annotation, the study applied Named Entities Recognition models, with a preference for the Bert pre-training vector model. Simultaneously, another research investigation, "An Investigation of Skill Requirements in Artificial Intelligence and Machine Learning Job Advertisements" by Amit, V. and Kamal, L. (2022), focused on skill prerequisites in AI and machine learning job ads across the USA. Employing content analysis on data from Indeed.com, this study proposed an alternative view of employers' expectations and highlighted crucial skills based on their relative frequency. Both studies contribute valuable perspectives for projects aiming to assist individuals in understanding and cultivating essential skills for success in specific fields, using techniques such as Python programming and Natural Language Processing (NLP). In my project, inspired by these methodologies, I aim to identify required data analyst skills while adding an industry-specific dimension through comparative analysis across different sectors.

3. Objectives and expected contributions

The primary objectives of this project are to facilitate individuals in discerning the pivotal skills requisite for a data analyst role and to offer comprehensive insights into the skill sets sought after by various industries. Employing the Python programming language, coupled with Natural Language Processing (NLP) techniques, we aspire to conduct a rigorous analysis of job descriptions pertinent to data analyst positions.

4. Methodology

To achieve the outlined objectives, the project will employ a multi-faceted methodology. Initially, web scraping will be conducted using Python libraries, specifically BeautifulSoup and Scrapy, to extract pertinent data from widely recognized job websites featuring data analyst roles. Subsequently, Natural Language Processing (NLP) techniques, including tokenization and sentiment analysis, will be applied systematically to process and comprehend the content within job descriptions.
Following the data collection and NLP phases, the project will undertake a thorough frequency analysis. This involves a meticulous examination of the frequency and prominence of specific skills identified within the compiled job descriptions, with the aim of pinpointing the most recurrent skills sought by employers.

5. Brief Project Schedule

Task	Detail	Assignee
1) Data Collection	Collect job descriptions from the website (Web scraping)	Thanapoom
2) Data Preparation	Apply NLP techniques to convert the text data into the suitable format for analysis (Tokenization)	Thanapoom
3) Feature Definition & Feature Engineering	Define a skill set that will be analyzed / Skill clustering to group similar skills into clusters	Thanapoom
4) Exploratory Data Analysis (EDA)	Identify the most recurrent skills sought by employees	Thanapoom
5) Data Interpretation	Interpret the results to display meaningful insights	Thanapoom
6) Documentation and Reporting	Create a summary report for the entire pipeline	Thanapoom

References

- Cao, L., & Zhang, J. (2021). Skill Requirements Analysis for Data Analysts Based on Named Entities Recognition. In *2nd International Conference on Big Data and Informatization Education (ICBDIE)* (pp. 64-68). Hangzhou, China. DOI: 10.1109/ICBDIE52740.2021.00023.
- Verma, A., Lamsal, K., & Verma, P. (2022). An investigation of skill requirements in artificial intelligence and machine learning job advertisements. *Industry and Higher Education, 36*(1), 63-73. https://doi.org/10.1177/0950422221990990