Project name:Job Market Analysis

SDLC (Software Development Life Cycle)

The Software Development Life Cycle is a systematic process for building software that ensures the quality and correctness of the software built. SDLC process aims to produce high quality software which meets customer expectations.The software development should be completed within the pre-defined time frame and cost.

SDLC Phases

The entire SDLC process is divided into the following stages

- Phase 1: Requirement collection and analysis

- Phase 2: Feasibility study

- Phase 3: Design

- Phase 4: Coding

- Phase 5: Testing

- Phase 6: Installation/Deployment

- Phase 7: Maintenance

Platform Knowledge

What is data science?

Data science is the study of data to extract meaningful insights for business.It is a multi disciplinary approach that combines principles and practices from the fields of mathematics, statistics, artificial intelligence and computer engineering to analyze large amounts of data. This analysis helps data scientists to ask and answer questions like what happened, why it happened, what will happen, and what can be done with the results.

Importance of data science

Solves real problem using data.Modern organizations have lots of data.Online systems and payment portals capture more data in the fields of e-commerce, medicine, finance, and every other aspect of human life. We can process the information and predict the results.

Objective:

To find Companies most probable to hire an ML Engineer/Data Analyst Applicant in respect to his/her skillset.

To analyse Machine Learning Job Market in India with respect to the given problem statement using Segmentation analysis and outline the segments most optimal to apply or prepare for Machine Learning Jobs.

Attribute Information:

'Job.Title' : Job title

'Job.Description': Job description

'Rating': Job rating

'Company.Name': Company name

'Location': Location of the company

'Headquarters': Company headquarters

'Size': the Size of the company

'Founded': The year it was founded

'Type.of.ownership': Public or private ownership

'Industry': Type of industry company belongs

'Sector': Type of sector company belongs

'Revenue': Yearly revenue by company

'Competitors': Company competitors

'Python': Does description has python keyword in it 1 if yes 0 if no

'R.Prog': Does description has R prog keyword in it 1 if yes 0 if no

'Excel': Does description has Excel keyword in it 1 if yes 0 if no

'Hadoop': Does description has Hadoop keyword in it 1 if yes 0 if no

'SQL': Does description has SQL keyword in it 1 if yes 0 if no

'SAS: Does description has SAS keyword in it 1 if yes 0 if no

Preprocess the data:

Remove the irrelevant features.Change the data types of the columns and perform null value treatment.Convert formatted data column to date time format.When preparing data to include in your reports, it's important to ensure that the information is organized.Performed null value treatment,data transformation and Outlier treatment on the columns of the dataframe. Both Industry and Location columns have -1 value which has been replaced by mode value in respective columns using str.replace().

EXPLORATORY DATA ANALYSIS:

Performed Univariate,Bi variate and multi variate analysis on the data. I had plotted a countplot which gives the following conclusions.Number of data scientist job vacancies in Sanofi is 12 Number of Data Analyst job vacancies in citi is 12. Python is top most demanded skill is requred for data scientist job. For data Analyst Excel, python are most demanded skills required . For Machine Learning Engineer Most demanded skill is Python For Data Engineer Most demanded skills are Sql, python and excel. Data Scientist,Data Analyst,Data Enginner , Machine Learning Engineer job titles are interchangeably used for search AI jobs.Data Scientist , Data Analyst are most frequent job titles used.

Hardware Requirements:

- CPU: A modern multicore processor, such as an Intel Core i5 or i7, is recommended for training and deploying machine learning models.
- RAM: A minimum of 8GB of RAM is recommended, although more may be required for larger datasets.
- GPU: A powerful graphics processing unit (GPU) can significantly speed up training and inference of deep learning models. Nvidia GPUs are the most commonly used for machine learning tasks.
- Storage: A large amount of storage is required to store and process large datasets. Solid-state drives (SSDs) are recommended for fast access to data.

Software Requirements:
- Operating System: Most machine learning frameworks are compatible with popular operating systems such as Windows, macOS, and Linux.
- • Python: Python is the most commonly used programming language for machine learning, and most machine learning frameworks are built using Python.
- • Machine Learning Frameworks: There are many machine learning frameworks available, such as TensorFlow, PyTorch, Keras, and scikit-learn.
- These frameworks provide tools and APIs for building, training, and deploying machine learning models.
- • Development Tools: IDEs such as PyCharm, Jupyter Notebook, and Visual Studio Code are commonly used for machine learning development.
- • Libraries: Libraries such as NumPy, Pandas, and Matplotlib are commonly used for data manipulation, analysis, and visualization.
- In summary, to get started with your project, you will need a computer with a powerful CPU and GPU, a minimum of 8GB of RAM, and large storage capacity. You will also need to install Python and various machine learning libraries and frameworks.

Existing System:
The existing system was able to analyse the skills required for data science jobs.We visualized the number of job vacancies job title wise and city wise. There are 258 Vacancies for Data Scientist roles in Bengaluru There are 87 Vacancies for Data Scientist roles in Mumbai There are 45 Vacancies for Data Scientist roles in Hyderabad.

Proposed system:
We will group the data points into clusters and analyse the job details in each clusters.

Conclusion:

Q.1) Find Companies most probable to hire an ML Engineer/Data Analyst Applicant in respect to his/her skillset.

Ans. Top Most Companies which are hiring are
ZoomRx','Amazon','EY','Walmart','Citi','Sanfoi','Matelabs Innovations Pvt. Ltd. etc.We visualized

Q.2) To analyse Machine Learning/Data Analyst Job Market in India and outline the segments most optimal to apply or prepare for Data Analyst/ Machine Learning Jobs.

Ans.Top Most Locations Which Are Hiring Machine Learning/Data Analyst in India Are Beangaluru,Mumbai,Pune,Chennai,Hyderbad.Top Most Demanded Skills for prepare Data Analyst/ Machine Learning Jobs are Python,Excel,SQL.