**Yash Agrawal**

**🕿:+919860214804**

**🖂:yrayash@gmail.com**

|  |
| --- |
| **Career Objective** |

To apply analytical, problem-solving skills on challenging career assignments in data science field. Passionate to learn new things and use learned things to develop data-driven creative solutions.

|  |
| --- |
| **Professional Summary** |

* **1.5+** years of experience in Data Science field and strong knowledge in **Programming, Statistics** andbuilding **Predictive models**.
* Main areas of expertise are **data wrangling**, building models and weaving a story around the findings.
* Experience in **Python** libraries like **Numpy**, **Pandas**, **Matplotlib** etc.
* Have sound knowledge about **NLP**, **Deep Learning** which includes **CNN,** **Recurrent neural network**.
* Experience in Data Loading, Web Scraping, Storage, and dealing with File Formats like CSV, Excel, JSON, XML, Binary Data.
* Interacted with **SQL**-based relational and **NoSQL** databases dependent on the performance, data integrity, and scalability needs of an application.
* Good Knowledge about how to interact with HTML and Web APIs for Web Scraping.
* Extensively worked on data preparation stages like loading, cleaning, transforming, and rearranging data.
* Have a good knowhow about merging and combining datasets using **Pandas**.
* Experience in **Data Transformation** which includes filtering, cleaning, rearranging data.
* Have a good knowhow about validating the data using **EDA** Techniques: Central Tendency, Dispersions, Quartiles/Percentiles, Standardization
* Performed various operations like removing duplicates, Transforming Data Using a Function or Mapping, Replacing Values, Discretization and Binning, Detecting and Filtering Outliers, Permutation and Random Sampling, Computing Dummy Variables under Data Transformation.
* Experience in data munging for string and text processing using **Python**.
* Extensively worked in area of plotting and visualization using various Python libraries like Matplotlib and Seaborn
* Experience in several Python libraries like **Scikit-Learn** that provide solid implementations of a range of machine learning algorithms
* Good knowledge on **PCA** for Dimensional Reduction for **Supervised data**, Cluster & Segmentation Analysis for **Unsupervised data**

|  |
| --- |
| **Technical Skills** |

Statistical Skills : Descriptive and inferential Statistics, EDA, Test of Hypothesis, Model Validation

And diagnostics

Machine Learning Skills : Regression analysis, Logistic Regression, Decision tree, Ensemble Models

Recommender system, Anomaly detectionPCA, Cluster analysis, Gradient decent, SVM, Neural Network, Time series, NLP

Languages/Tools Skills : Python, R, TensorFlow, SQL, Tableau, Git

Operating Systems : Windows 10/9x/XP, UNIX.

|  |
| --- |
| **Educational Qualifications** |

* **B.E** in **Information technology** from Pune Institute of Computer Technology affiliated to University of Pune in 2018 with **First Class**.
* **HSC** from Govt. Technical Junior College, Akola (Maharashtra State Board) in 2014 with **85.38%**.
* **SSC** from Seth Bansidhar High School, Telhara (Maharashtra State Board) in 2012 with **91.43%**.

|  |
| --- |
| **Work Experience** |

**Hydes Solutions Ltd., Hyderabad Jr. Data Scientist**

**Project** : EHA

**Client**  : P&W

**Duration** : July’ 19 - .

**Description:**

The project of Pratt & Whitney is a predictive maintenance engine health analytics project. It had an objective to enhance the maintenance operations and planning of time-based preventive maintenance by applying data science techniques and machine learning algorithms for predicting more accurate maintenance requirements. Observing engine's health and condition through sensors and telemetry data is assumed to facilitate this type of maintenance by predicting Time-To-Failure (TTF) or Remaining Useful Life (RUL) of in-service equipment.

**Responsibilities:**

* Involved in data preparation stage with activities like loading the data from source file into **Pandas** dataframe, assign columns names, remove empty or unwanted columns and merging with the help of Pandas.
* Explored the loaded data by running some Pandas dataframe methods like **.describe(), .head(), .tail(), .dtypes** to get more information about the dataset.
* Checked for null (NaN) values by running Pandas dataframe method **.null(), .sum()**. Accordingly, some columns were excluded from the data as they had high number of missing values.
* Performed outlier detection by running Pandas dataframe **.quantile()** method, SciPy **stats.zscore**, or manually by identifying the data points above or below [mean ± n standard deviation], where n can take the value 2 or 3 assuming normal distribution.
* Merged dataframes based on index or common key. For example, labels for test data were in

separate source data file. This was done by using Pandas **.concat() and .merge()** methods.

* Performed **EDA** on the dataset which includes exploring each feature individually and inspecting data using Groupby function on various features.
* Created **Regression** and **classification** labels for training data.
* **Feature extraction** is also applied to the training and test data by introducing additional two columns for each of the 21 sensor columns: rolling mean and rolling **standard deviation**
* EDA performed using various data visualization python library like **Matplotlib**, **Seaborn**.
* Feature variability, distribution,and correlation were examined to uncover underlying structure and extract important variables.
* Features with high variability were checked for correlation with other features and regression label.
* Scatter matrix was also used to check the distribution and correlation of features
* Prepared a number of EDA charts to have more insights on each feature individually.
* The following machine learning algorithms were tried and their performance metrics were calculated and evaluated: **Linear Regression**, **LASSO Regression, Ridge Regression, Decision Tree Regression, Polynomial Regression, and Random Forests Regression**.
* Calculated key regression evaluation metrics like **Root Mean Squared Error** **(RMSE),** **R-squared (R2), Mean Absolute Error**.
* Tried answering three essential questions in predictive maintenance: When an engine will fail? Which engines will fail in this period? How better could maintenance be scheduled?.

**Hydes Solutions Ltd., Hyderabad Jr. Data Scientist**

**Project**  : GrofersMarket Basket Analysis

**Client** : Grofers India Private Limited.

**Duration** : Dec’ 18 – May’ 19.

**Description:**

Market basket system has developed for Grofers India Pvt. Ltd., which is a online grocery platform. Grofers market basket system (GMB) collects data from various departments of various item categories. GMB platform analyzes various attributes and summarizes the dataset for better understanding. System is designed to report which products are often shopped together. It will help understand the customer preference in terms of products they buy collectively.

**Responsibilities:**

* Involved in data preparation stage with activities like loading the data from source file into **Pandas** dataframe, assign columns names, remove empty or unwanted columns and merging with the help of Pandas.
* Explored the loaded data by running some Pandas dataframe methods like **.describe(), .head(), .tail(), .dtypes** to get more information about the dataset.
* Checked for null (NaN) values by running Pandas dataframe method **.null(), .sum()**
* Implemented data transformations which includes flitering, cleaning, rearranging the data.
* Joined dataframes using Pandas .**concat()** and **.merge()** methods.
* Performed **EDA** on the dataset which includes exploring each feature individually and inspecting data using Groupby function on various features.
* EDA performed using various data visualization python library like **Matplotlib, Seaborn.**
* Converted order data into format expected by the association rules function and displayed summary statistics for order data.
* Performed association analysis in **Python** using the **apriori algorithm**.
* Calculated and recalculated item **frequency and support**.
* Created table of **association rules** and computed relevant metrics.
* Formulated the assocaition rules with their corresponding **support, confidence and lift**.
* Once item pairs have been identified as having positive relationship, recommendations can be made to customers in order to increase sales.

**Hydes Solutions Ltd., Hyderabad Jr. Data Scientist**

**Project** : Online Grocery Reviews Analysis

**Client**  : Grofers India Pvt. Ltd

**Duration** : July ‘18-Nov`18

**Description:**

Sentiment Analysis was performed on reviews related to online grocery platform, Grofers. System gathers and extracts publically available customer reviews regarding the company. It then performs various analytics, such as aspect based analysis and sentiment analysis to finally gauge the customer response and satisfaction.

**Responsibilities:**

* Collection of customer reviews from various online platforms using Python libraries like **Goose**, **BeautifulSoup**.
* Analyze the change requirement and adjust the system accordingly.
* Performed Pre-processing stage by removing stop words using python library NLTK.
* Parsed web pages and performed tokenization using **NLTK**.
* Analyzed presented data in the form of various graphs like bar graphs/ pie charts.
* Performed **sentiment analysis** using **nltk.sentiment.vader** and **SentimentIntensityAnalyzer**
* Provided help to the team about **crawlers**, **NLP** technologies.

|  |
| --- |
| **Key Achievements** |

* “Pride@P&W ” award from P&W for EHA Project.
* Published a paper entitled ” Implementing Social CRM System for an Online Grocery Shopping Platform Using Customer Reviews ” at International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue: 08 on Aug 2018. <https://www.irjet.net/archives/V5/i8/IRJET-V5I8193.pdf>
* Secured 1st position in a National Level Contest PRADNYA held amongst 120+ participants under Impetus and Concepts 2016 (Cash Prize of INR 15000)

|  |
| --- |
| **Personal Details** |

Date of Birth : 08-01-1997

Languages Known : English, Hindi, Marathi, Marwari.

Marital Status : Single

Passport Number : S8340573

Passport Validity : 04/09/2029

Permanent Address: Akola, Maharashtra

|  |
| --- |
| **Declaration** |

I hereby declare that above furnished information is true to my knowledge and I bear complete responsibility for the correctness of the above mentioned information.

Date:

Place: Hyderabad

(Yash Agrawal)