HARSHAL GARG

CONTACT

PHONE:
+91-9977816300

EMAIL:
harshalgargwork@gmail.com

LINKEDIN
linkedin.com/in/harshalgarg/

GITHUB:
github.com/gargharshal

MEDIUM:
harshalgarg.medium.com

EDUCATION

B.E. – Electrical and Electronics.
2014 – 2018
UIT- RGPV, Bhopal

SKILLS

Power BI Advanced, AWS certified

Machine Learning - Sklearn, MatplotLib, Seaborn

Deep Learning - TensorFlow, CNN, NLP

Python - Pandas, NumPy

Basics - Java, SQL, SAP HANA CVs, Matlab

Agile Methodology

PROFESSIONAL SUMARRY

- I have around 2.5 years of experience in IT working in the field of **Data and Analytics** with hands-on experience in technologies like **Python**, **Power BI**, **SQL**, **AWS** etc.
- Handled database, created reports, deploying project according to client's requirement while following **Agile** methodology.
- Self-taught **Data Science** and **machine learning** enthusiast with hands-on experience from projects.

WORK EXPERIENCE

Infosys Limited - Senior Systems Engineer
JAN 2021 – Current

Data Migration and Report Recreation (ongoing)

- Migration of database from On-premise **SAP BW** to **AWS Redshift** while following **Agile** Methodology.
- I decided on **architecture**, created **mapping docs** of the new vs old databases' attributes created tables and ingested the data.
- I decided on architecture, built, and deployed reports on **Power BI** and connected them to the new database.
- I worked on **AWS Glue Job** with **python** to generate required outputs like HTML files by ingesting data from **AWS S3** and **Redshift**.

Infosys Limited - Systems Engineer
APRIL 2019 – DEC 2020

BW Monitoring and Support

- I monitored the process chains of 4 systems in **SAP BW**.
- I had to reschedule data transfers jobs based on time, load, or any other business requirement.
- I used to monitor ticket in Service Now

Data Migration and Report Creation

- Migration of database from On-premise SAP BW to SAP HANA.
- I designed and built Graphical Calculation Views in **SAP HANA.**
- I created the reports on **Power BI** which included making a **data model**, different type of **visuals**, **DAX Queries**, **time intelligence function** and deploying the reports on Power BI web.

Infosys Limited - Systems Engineer Trainee
NOV 2018 – APRIL 2019

Completed Infosys Foundation Program

- Generic Training - Python, SQL
- Stream Training - Power BI, MongoDB, Java, pig, hive, HBase, MSSQL

CERTIFICATION

Deep Learning - Coursera by DeepLearning.ai – Credentials

Stanford Machine Learning - Coursera by Andrew Ng – Credentials

AWS Certified Cloud Practitioner – Credentials

Matlab Onramp – Credentials

PROJECTS

Dog Breed Identification – CNN (link)

- Using Kaggle dataset to create a dog breed identifier based on an image.
- I used **Transfer Learning** by using a **pretrained Keras model**.
- I trained every model on a small dataset to find the best one and then **NASNetMobile** model was chosen based on **accuracy** and **parameter count**.
- I created **callbacks** to **create logs** and to prevent the model from **overfitting.**
- The model was **trained** on the full dataset and saved as a **'.h5'** file which was used to make **predictions**.

Bulldozer Price Prediction - Random Forest (link)

- Given the data of previous sales prices of bulldozers, we are trying to predict the sales price of the bulldozers which have similar characteristics.
- I first analyzed the data to find **missing values** and the **most important features**.
- I filled the missing values and made compatible the data was for modelling.
- I choose **RandomForestRegressor** based on Sklearn Algorithm Cheat Sheet
- After hyperparameter tuning, I finalized the model, trained and tested it.

Heart Disease - Logistic Regression (link)

- Using Kaggle Dataset to predict if a person has heart disease or not.
- The data on Multiple models to find the best one based on **precision**, **recall**, **f1-score**
- Understanding the parameters involved and their importance using **histograms**, **scatter plots**, **Confusion Matrix**.
- Using **Logistic Regression** as the final model, fine tunning hyper parameters with **RandomizedSearchCV** and then training the best model.
- Understanding the model performance using roc curve, classification_report, etc.