Praveen Reddy Nelli

Data Scientist at CGI

Email: reddypraveen722@gmail.com

Mobile: +91-8919098852

LinkedIn URL: https://www.linkedin.com/in/Praveenreddi722/

Professional Summary:

- Data Scientist with 4.6 years of industry experience in Statistical Modelling, Predictive Modelling algorithms to solve challenging business problems.
- Experience in mapping business requirements to Data Science solutions and having experience implementing POC's.
- Expertise in Machine learning, Deep learning , Computer vision models and having experience deploying scalable and serviceable AI models using RESTful services and Docker.

Work Experience:

Project1:

Domain	:	Travel
Client	:	LUFTHANSA
Task	:	Predictive Failure Analysis and Anomaly detection on Time series
Algorithms	:	Vector Auto Regression (VAR), Facebook Prophet, Isolation Forest

Responsibilities:

- Gathered Multivariate Time Series data collected from ElasticSearch database.
- Queried the database using elasticsearch-dsl and cleaned it using Python for modeling.
- Performed EDA: Data integration, outlier analysis, missing value treatment, correlation analysis.
- Conducted statistical checks: Causality test, Cointegration test, Stationarity check.
- Applied Multivariate-forecasting algorithm on historic data to predict the future outcomes of the system metrics and evaluated the forecasts using RMSE (Root Mean Squared Error), Mean Absolute Percentage Error (MAPE) etc.
- Applied unsupervised machine learning model on system metrics to identify the anomalies present in the data.
- Deployed the models using Flask and the results obtained are sent to the customized monitoring tool (UI) via Rest API.

Project2:

Domain	:	Retail
Client	:	SODEXO
Task	:	Document Verification for KYC with OCR and Computer Vision
Algorithms	:	Faster R-CNN, Tesseract-OCR, Google Vision-OCR, Verhoeff

Responsibilities:

- Pre-processed the input images: Image resizing, Image denoising (Gaussian Noise filter) etc.
- Annotated a set of images with different classes using LabelImg tool for feature extraction.
- Trained an Object Detection model (Faster RCNN) chosen from TensorFlow Object Detection API on custom data set using pre-trained weights.
- Exported the Inference Graph and performed Object detection to identify the ROI (Region of interest).
- Performed OCR on the Region of interest and extracted the text.
- Validated the extracted text with regex and verhoeff algorithm, the result is then sent via Rest API in Json format.

Project3:

Domain	:	Travel
Client	:	LUFTHANSA
Task	:	Sentiment Analysis on Twitter Streaming data
Technology	:	Python, TwitterAPI, TextBlob, Flask, Streamlit, Heroku

Responsibilities:

- Established the connection using keys and tokens to authenticate Twitter API.
- Extracted the data by querying Twitter API, which fetched the data with information such as username of the person, likes for the tweet, location of the user and stored the data in a dataframe.
- Preprocessed and cleaned the data by removing links, special characters using regex statements.
- Analyzed the sentiment of the data using TextBlob which gives polarity of the sentence ranging from [-1,1].
- Plotted Pie charts and bar charts for different sentiments and generated WordCloud for positive and negative sentiments.
- Created an API using streamlit and flask and deployed the project on Heroku.

Project4:

Domain : TRAVEL
Client : Lufthansa
Task : Perform log analysis on different application servers.
Tools : Linux, Elastic Search, Logstash, Kibana, Google Cloud

Responsibilities:

- Performed log analysis and parsed the logs to logstash from different application servers.
- Integrated the logs into Elastic search database using file beats.
- Fetched the logs from Elastic search and displayed the results in Kibana dashboard, which reduced the total manual effort of 60%.
- Installed Metric beats to collect and display the server related metrics in Kibana.
- Implemented several automation scripts using shell for KPI's alerts based on their SLA's.

Personal Projects:

Project 1: Face Mask Detector

- Collected and annotated a set of 2K images with two different class labels i.e. mask and No mask
- Generated TFRecords for the training of the object detector.
- Configured the training files and trained the model (SSD_mobilenet) chosen from TFOD API using pre-trained weights.
- Exported the inference graph and performed object detection to test the model.

Project 2: Question pair similarity problem

Task : Given pair of questions to predict if they are duplicates or not.
Algorithms : XGBoost, Logistic Regression, Support vector machine (SVM)

Responsibilities:

- Loaded the data from the csv files and stored in a dataframe.
- Performed exploratory data analysis (EDA): Null or missing values, distribution of final class labels etc.
- Performed basic feature extraction on the data and applied preprocessing steps to remove html tags, punctuations, stopwords etc.
- Converted the text data into vectors using Bag of words (BOW) and TF-IDF and scaled the data using normalization and standardization techniques.
- Split the data into train and test set, applied various machine learning models on the training data, and evaluated the models on test set.
- Evaluated the results of the misclassification using log-loss and binary confusion matrix methods.

Project 3: Text classification on food reviews

Task : Given a review, determine whether the review is positive or negative
Algorithms : KNN, Naïve Bayes, Logistic Regression, Decision trees.

Responsibilities:

- Loaded the data from the csv files and stored in a dataframe.
- Performed exploratory data analysis and removed duplicates, filled missing values and analyzed the target variable.
- Applied preprocessing steps on the text data to remove html tags, punctuations, stopwords etc.
- Converted the text data into vectors using Bag of words (BOW) and TF-IDF and scaled the data using normalization and standardization techniques.

- Split the data into train and test set, applied various machine learning models on the training data, and evaluated the models on test set.
- Evaluated the results of the misclassification using AUC and binary confusion matrix methods.
- Deployed the model using Flask framework and the response is sent via RestAPI in json format.

Achievements:

- Received Star award for recognizing as best employee of the year in 2019.
- Performance and Outstanding Contributor award as per managers based rating.

Certifications:

- Machine Learning with Python(ML0101ENv3) Issued by IBM.
- Deep learning masters (DL012007) issued by iNeuron.
- Data Analysis with Python (DA0101EN) Issued by IBM.
- Neural Networks and Deep learning issued by coursera from deeplearning.ai

Education:

- Bachelor of Technology from JNTU College of engineering Anantapur in the year 2016 with 77%.
- Intermediate from Narayana Junior College in the year 2012 with 96%.
- SSC from Sri Srija School in the year 2010 with 91.3%.

Skills:

Algorithms	:	Machine Learning, Deep Learning, Computer Vision
Programming Languages	:	Python, Shell
Visualization	:	Kibana, Matplotlib library
Other libraries	:	numpy, pandas, Scikitlearn, keras,Tensorflow,Statsmodels
Framework	:	Flask.

Personal Details:

Date of Birth	:	May 15,1995
Nationality	:	Indian
Father's Name	:	Mr. Satya Narayana Reddy Nelli
Languages Known	:	English, Hindi, Telugu

Declaration:

I hereby acknowledge that all the above given details are true up to my knowledge and belief.

(Praveen Reddy)