R Sankeerthan Reddy

E-Mail id: sankeerthanratna.12@gmail.com Mobile: +91 8019839337

Professional summary:

- A proactive fast learning individual having 2+ years of experience IT industry
- Currently working as a **Data Scientist** with **Moonstone Info Tech pvt ltd**.
- Skilled at team collaboration while working independently in remote environments
- Skilled in:
  - Rich research analysis skill-set on Business analysis, data analysis, statistical analysis & modeling on implementing data-driven business processes with risk identification through decision models by leveraging skills in Machine Learning Algorithms, Data/Text Mining techniques, Data Visualization, etc.
  - Experience in performing Descriptive, Predictive, and Prescriptive Analytics by employing various data mining techniques like statistical analysis, hypothesis testing, regression analysis, unsupervised & supervised machine learning algorithms, advanced analytical techniques, text mining with sentiment analysis, and NLP
  - Experienced in identifying data patterns, trends, analyzing data to develop best predictive models for Processing, cleansing, and verifying the integrity of data used for analysis
  - Good exposure to data mining techniques like clustering, developing association rules
  - Hands-on experience in tools like R, Python, Spark
  - Experienced in developing reports and dashboards using BI tools like Tableau and PowerBI
  - Strong Understanding of lifecycle of analytics projects
  - Values opportunities to stretch professionally and help others to grow and develop
  - Exposure towards Hadoop Ecosystem
  - Knowledge on Keras, Tensor Flow for Deep Learning including CNN, RNN, RCNN and GANS.

Tool skills

Programming	R, Python, Spark, Databricks
Statistical	Minitab, XL Miner, Microsoft Excel
Reporting	Tableau, Power BI, MS Excel
Database	MySQL, MSSQL, HDFS

Professional Experience

Data Scientist – Moonstone info tech pvt ltd **Sep' 2018 -- Till date**

Education Details:

- Bachelor of Technology – Mechanical, 2018, Sri indu college of engineering and technology, Hyderabad.
- Board of Intermediate Education – Math, Physics and Chemistry, 2014, Hyderabad
- Board of Secondary School Education – 2012, Bhupallpelly, TS.

Projects

1. Presently working on **Site code Prediction** project using **Python-- Ford**

Objective: Create a Machine Learning and Artificial Intelligence based Application to detect the entered GSDB site code is correct or incorrect and then suggest the correct one if entered site code is incorrect along with probability.

Responsibilities

- Exploratory Data Analysis to understand the distribution and different Cluster Segments on basis of Supplier codes
- Handling missing values using different imputation methods
- Handling one hot encoding to convert the categorical data to numerical
- Functionality includes the classification of Supplier code based on the given Location, Base, prefix values and suffix values
- Predicting Actual Supplier code for future by training historical data using machine learning methods (Neural Networks, Naïve Bayes)
- Optimized the models using hyper parameter tuning
- Handling over fitting and under fitting of the models using regularization techniques
- Implemented Authorization for data security in between Java and Python
- Deployed the application in Cloud environment
- Visualized the data in multiple ways using tableau

Tools & Packages: Python, MySQL, Python Packages: sklearn, Numpy, Pandas, Flask Frame work

Data Pipeline

- Data is stored in SQL database in multiple tables
- Written SQL queries and brought the data together into one table
- Data is extracted using the pyodbc
- After performing the model building and predictions the results are stored in mongo DB using pymongo

2. Insurance Claim processing – The General

Objective: Processing the claims over online without manual intervention and capturing the requests over various modules of data (audio, doc, pdf and images) using NLP and RE, Identifying the damage location and severity using Deep learning approaches

Analysis

- Received Various forms of data includes audio files, doc files, pdf's and Images
- Able to extract the required fields of the customer information and claim related information using RE and NLP
- Using tesseract and google speech recognition able to extract the text from images and audio files
- Process the attached image of damaged vehicle
- Able to predict the location and severity of the damaged part using Deep learning methods such as CNN using Tensor flow
- Developed a user interface to upload the customer supporting documents

Responsibilities

- Written python query to separate the types of data receiving
- Applied textract, tesseract and google speech recognizer to process the text
- Using NLP techniques identified the required fields of information such as area and Names
- Using regular expressions extracted the required other information like claim number, SSN ID and VIN number etc.
- Verifying the damage severity of vehicle using CNN with the help of transfer learning

- Used VGG pre-trained network
- Once the severity is detected the captured data will be passed to the database
- Implemented regularization to overcome the over fitting and under fitting issues

Tools: Python, Mongo dB

Data Pipeline: Once we receive the data which is verified will be passed to mongo dB Pymongo

3. **Text classification using NLP and spacy**

Objective: The goal is to identify the sentences which are speaking negative or positive using the words with in the sentence and classify to which group that sentence belongs among multiple groups as per the requirement

Analysis

Extracting the raw text from pdf using OCR engine and separate each sentence with in the paragraph and perform sentiment analysis in each statement using NLTK (POS tagging, key phrase extraction and word vectorising) then perform machine learning algorithms like tree classifier in python and classified to which group that each word belongs to and visualizing them

Responsibilities

- Extracted the raw text from pdf's using tesseract
- Separated the sentence using NLP
- Perform the sentimental analysis using text blob
- Performed POS tagging, word2vec
- Perform machine learning algorithm to classify the group

Tools & Techniques: Python, R, Random Forest, Decision tree,

Technical skills

Regression Analysis	Linear Regression, Logistic Regression, Multinomial Regression, Poisson Regression, Offset Regression, Negative Binomial Regression, Zero Inflated Regression, Hurdle Regression
Supervised Machine Learning Algorithms	Decision Tree, Random Forest, KNN, SVM, Naïve Bayes, Neural Network
Unsupervised Machine Learning Algorithms	Cluster Analysis, PCA, SVD, Market Basket Analysis, Recommendation Systems, Latent Dirichlet Allocation
Times Series Analysis	Forecasting (Model Based & Data Driven)
Text Analytics	Text Mining, Sentimental Analysis & Topic Extraction using NLP (LDA)
Advanced Analytics	Feature Engineering, Imbalanced Classification, Boosting, Bagging, XGBM, Gradient Descent, Deep Learning (MLP, CNN, RNN)
Statistical Analysis Techniques	Confidence Intervals, Hypothesis Testing

Declaration:

I hereby declare that the information furnished above is true to the best of my knowledge and belief.

(R Sankeerthan Reddy)