Name: Abdulkadar Babbar
Email: abdulkadarbabbar@gmail.com
Mob No: +919082425027
City: Pune, India
LinkdinID: https://www.linkedin.com/in/abdulkadar-b-085254129

Summary

- Professional Data Scientist with 5+ years' of experience in core Data Science and Analytics projects
- Post graduate in data science, business analytic and big data
- Strong Hands on experience in programming language python, R and SQL

Skill Highlights

Analytical & Data Science Skills	**Descriptive analysis, Data visualisation, Predictive modelling, Time Series forecast model, Supervised and Unsupervised ML model, Natural Language Processing(NLP)**
Data Engineering Skills	**ETL pipeline, Orchestration with Apache Airflow, Web scrapping, Model deployments (MLOPs) & Docker**
Tools and Language	**Python, R, SQL, MATLAB, Tableau, Qlik(basic)**
Hands on Cloud Platforms	**AWS (Lambda, EC2, Redshift, Athena, Glue, ECR, S3) & Azure (Basic)**

Work Experience

Nagarro PVT LTD, Pune, India

Associate Staff Engineer/ Lead Data Scientist **Aug 2021 – Current**

- Working as a Data Scientist for client in order to improve their decision process in Manufacturing & Supply Chain domain.
- Created and implemented forecasting models to help stakeholder in decision process
- Responsible for end to end development & deployment of project cycle
- Setup whole ETL architecture using airflow on production
- Also, Working with internal COE team of Nagarro for accelerator development in Machine learning

SG Analytics, Pune, India

Data Scientist **Aug 2019 - Aug 2021**

- Worked as Data Scientist for client in media domain, especially for their OTT platform app
- Responsible for development of ETL & Machine learning model pipeline and delivered as a package using containerize method such as Docker

Name: Abdulkadar Babbar
Email: abdulkadarbabbar@gmail.com
Mob No: +919082425027
City: Pune, India
LinkdinID: https://www.linkedin.com/in/abdulkadar-b-085254129

Ernst & Young LLP, Bangalore, India

Data Scientist **Oct 2018 – July 2019**

- Worked as Data Scientist for Finance & Audit , HR team to optimize their existing process and make it more efficient

DigiKredit Finance Pvt Ltd, Mumbai, India

Data Scientist **Sep 2017 – Oct 2018**

- Worked as Data Scientist for core credit and operation team in lending business
- Responsible for development of predictive model and make model explainable to credit manager and operation manager to improve their decision

Edelweiss Finance Service

Data Science Intern **June 2017 – Aug 2018**

- Worked as Intern for IPO team to provide insightful insight for IPO broker and predict broker performance using Regression analysis

Education

- **Post Graduate Program** in Data science, Business Analytics and Big Data from Aegis School of Business, Data Science and Telecommunication jointly certified with IBM, Mumbai

Sep 2016 –July 2017

- **Bachelor of Engineering** in Electronics and Tele-Communications from D.Y. Patil College of Engineering, Pune **July 2011 –July 2015**

Project From Past Work

- **Customer Review Classification [Python ,Docker and AWS(Batch)]**
    - Perform transformation as cleaning, formatting & filtering of large volume of data includes unstructured text getting from third party vendor's app using python and load it into the redshift table.

Name: Abdulkadar Babbar
Email: abdulkadarbabbar@gmail.com
Mob No: +919082425027
City: Pune, India
LinkdinID: https://www.linkedin.com/in/abdulkadar-b-085254129

- Built a Natural Language Understanding based system to automatically categorize customer reviews into pre-defined categories and generate visualizations around the data for stakeholder engagement
- Containerize entire project using Docker to run it on cloud / client environment
- Automate entire process on AWS cloud by running Docker image as per schedule on AWS batch to train model as well as to predict new reviews as per scheduler.
- This Auto ML architecture generates report delivery with minimal human interaction

- **Customer Banking Analysis Application Software [Python]**

  - Developed a tool that perform analysis on customer bank statement and Classify transaction type by looking into transaction description(text).
  - For XML PDF, developed XML scrappers to get data in structures format and load in data base, it has almost 90 % of coverage
  - For OCR, third party API used to get dump in CSV
  - Perform different descriptive and predictive analysis that generate summary of customer banking
  - This tool reduces human error and eta that was there when banking generates by process associates.
  - Result includes complete banking summary, end of day balance, Fraud Transaction and Transaction Classification

- **Skip Level Meeting Analysis [Python-Topic Modelling]**

  - Skip level meeting has feedback data in the form of text
  - Tool extracted important topics from huge collected feedback using ML algorithm LDA
  - Model result helps upper management to get understand most repeated feedback & get prioritized issues.

- **Credit Mining Platform [Python]**

  - Developed web scraping scripts on given URL required for extracting information for customer credit and transform data to required format
  - Design Data pipeline workflow for hierarchical analysis to generate credit score.
  - Analysis includes sentiment score using NLP on text gathering from scrapper as well as generates score using financial or demographic variable by ML approach.

Name: Abdulkadar Babbar
Email: abdulkadarbabbar@gmail.com
Mob No: +919082425027
City: Pune, India
LinkdinID: https://www.linkedin.com/in/abdulkadar-b-085254129

- **Recommendation Engine to provide recommendation of show's for user (python):**
  - Developed recommendation engine by using combination of two filtering approach as:

    a. Collaborative Filtering (Restricted Boltzmann's Machine)

    i. RBM model is trained using Neural network with rating data
    ii. Recommendation is provided by reconstructing weights of Neural network for input user

    b. Content Base Filtering (Clustering – Shows & User's)

    i. Make a Clusters of shows using show attributes as "show details", "taxonomy", "tag" etc. (Clustering Algorithm - BIRCH)
    ii. Make a Clusters of USER'S using data that provides demographic & geographic information of User (Clustering Algorithm – KMEANS)
    iii. Provide user as input for which engine will provide recommendations shows using logic / set of rules
    iv. Content based filtering also useful for new shows since those have been very frequently watched and minimal rating so it possesses less likelihood of recommendation with collaborative filtering

  - Combined result from both above methods and drop the duplicates to get final recommendation from content base filtering. (**Result =a+b**)

- **Credit Score Card [R Studio and Tableau]:**
  - Developed score card using Machine Learning include roll rate and delinquency analysis
  - Data Source > Financial from ITR, Banking Data, CIBIL Data, Customer Demographic and Geographic Data, Alternate Data from Credit Mining APP.
  - The model is self-machine learning and use alternate data sources that makes it more dynamic and robust
  - Score card helps business at different step of credit process journey to take better decision