DATA SCIENTIST

ASHISH KUMAR

GET IN CONTACT

Mobile: 8335818761

Address: Ratu, Ranchi, Jharkhand-835222

Website : https://ashishkrb7.github.io/
Github : https://github.com/ashishkrb7
Linkedin : https://www.linkedin.com/in/ashishk766/
Gmail :ashish.krb7@gmail.com

PERSONAL PROFILE

A data scientist with hands-on experience in all levels of AI-ML-DL modelling, including performance, functional, integration, system, and user acceptance.

AREAS OF EXPERTISE

- Data Science
- Data analysis
- Data mining for Structured and unstructured dataset
- Simulation, modelling and optimization
- Hydrological modelling
- GIS modelling

TECHNICAL SKILLS

- Programming languages: Python, Matlab, R, C and Matlab
- Database operation: MSSQL-Server and MongoDB
- Data visualization: Power BI, Jupyter notebook

WORK EXPERIENCE

DATA SCIENTIST

CSS Corp Pvt. Ltd. (R & D), Hyderabad | 2019 – Present

- Translated business requirements into functional design
- Extensive data analysis of the company's data to identify the problem and solved using AI-ML
- Designed software and backend web service development
- Applied statistical techniques to diagnose the business problem

TEACHING ASSISTANT

IIT Madras | 2017 – 2019

- Trained students on various computer software
- Troubleshoot software problems
- Assignment preparation and correction

EDUCATION HISTORY

IIT MADRAS, CHENNAI

Master of Technology, 2019

- Degree in Hydraulics and Water Resource Engineering [Civil Engineering]
- Learned in-depth about how to optimize real life problem and process using technological systems
- MHRD GATE scholar

BIDHAN CHANDRA KRISHI VISHWAVIDYALAYA, WEST BENGAL

Bachelor of Technology (Hons.), 2017

- Earner degree on Agricultural Engineering with Gold medal
- Learned in-depth about how modern days technology can help the Agricultural sector
- NTS scholar (ICAR)

PROFESSIONAL SUMMARY

- Hands-on, successful and efficient Data Scientist with 1.9+ years of experience in Data Analysis, Data Science, Machine Learning, Deep Learning and data mining on structured and unstructured data, Data acquisition, Data Validation, Predictive modelling, Data Visualization, Web Scraping
- Skilled in performing Data Parsing, Data manipulation and Data preparation with methods including describing data contents, compute descriptive statistics of data, regex, split and combine, remap, merge, subset, reindex, melt and reshape
- Experience in using various packages in python like pandas, NumPy, SciPy, TensorFlow, PyTorch, Sklearn, PyCaret, Beautiful Soup etc
- Developed Dashboard and story point using Jupyter notebook and PowerBI. Developed various advanced interactive visualizations such as Heat map, Bubble chart, Treemap, and Line charts while working on motion chart, & Drill-down analysis
- Adept and deep understanding of Statistical Modeling, Multivariate Analysis, model testing, problem analysis, model comparison and validation.
- Involved in the entire data science project life cycle and actively involved in all the phases including data extraction, data cleaning, statistical modelling, and data visualization with large data sets of structured and unstructured data.
- Proficient in managing entire data science project life cycle and actively involved in all the phases of the project life cycle including data acquisition, data cleaning, features scaling, features engineering, statistical modelling (Decision Trees, Regression Models, Neural Networks, Support Vector Machine, Clustering, Bagging and boosting techniques), dimensionality reduction using Principal Component Analysis and Factor Analysis, testing and validation using ROC plot, K – fold cross-validation and Data Visualization etc.
- Worked and extracted data from various database sources like SQL Server, MongoDB, APIs and Web etc.

PROJECTS

Data analysis (June 2019-Oct 2019)
Role: Data Analyst
Summary: Done Exploratory data analysis (EDA) on different companies' data like Alcatel Lucent Enterprises, Blackboard, Family Dollar, Sophos and translated the business requirement into functional designs by building different classification, regressor and time series models
Requirements:

- Analyze and prepare dataset collected from Salesforce Customer relationship management(CRM) by applying historical models using python
- Perform data manipulation, data preparation, normalization and predicting model. Used Bagging technique (Random Forest) to predict SLA based on various features
- Used ARIMA model to predict the trend in tickets in coming months
- This project was to predict the SLA prior to ticket assignment which can assist the Agent in solving the ticket.
- Made PowerBI Dashboard with different line graphs, bar charts, word cloud with drill-down
- Sentiment and Urgency prediction by analyzing the textual data. VaderSentiment and Naive Bayes model were used for making the model.

Tools used: Python, Salesforce API, MS-Excel, Jupyter Notebook, PowerBI, ARIMA, Random Forest, Naive Bayes.

RNN-Chatbot (Oct 2019-Nov 2019)
Role: Data Scientist
Summary: Developed Retrieval type chatbot using encoder-decoder model using deep learning
Requirements:

- Research on trending Chatbot technology
- Data collection from the public domain
- Developed an encoder-decoder model using Long Short-Term Memory (LSTM) neuron
- Developed Restful API and hosted the app in an on-premise server

Tools used: Python, Flask, TensorFlow, Keras, MS-Excel, Jupyter Notebook, LSTM, NGINX, Docker, Anaconda, VScode

Persona-based smart case manager (Dec 2020-Mar 2020)

Role: Data Scientist

Summary: Develop a ticket routing Engine to route the incoming ticket to the best agent based on Agent persona and skill

Requirements:

- Participated in the development of application architecture and blueprint to define application component, platforms, interface and development tools
- Analyze and collect dataset from Salesforce CRM using Pandas, Numpy and request package
- Research on Persona requirement form prototype
- Developed package for Analytical Hierarchy process (AHP) and developed Flask app and hosted it on an on-premise server

Tools used: Python, Salesforce API, Jupyter Notebook, AHP, NGINX, Docker, VScode.

Agent Cobot-Digital Assistant(April 2020-May 2020)

Role: Data Scientist

Summary: Developed probable cause predictor using unsupervised machine learning technique on emails attached to tickets, in addition to analytics like sentiment, readability, subjectivity analysis, attachment information identifier and extractive summarizer

Requirements:

- Analyze and collect dataset from Salesforce CRM using Pandas, Numpy and request package
- Developed a Latent Dirichlet allocation (LDA) model using the Gensim package for identifying a topic in the email historical dataset and predict the topic in the incoming ticket
- Developed model for predicting the flags like sentiment, readability, attachment information and keywords and summary using Gensim and textstat package
- Developed Flask app and hosted it on an on-premise server

Tools used: Python, Gensim, textstat, Salesforce API, Jupyter Notebook, NGINX, VScode.

ALE Log Analytics(Jun 2020-Aug 2020)

Role: Data Scientist

Summary: Developing a recommendation engine to suggest the knowledge base based on the issue on Switch device logs. The client wanted to have AI-based log analytics tools for two of the devices, which generates a different variety of log.

Requirements:

- Participated in the development of application architecture and blueprint to define application component, platforms, interface and development tools.
- Collected data from Engineers and understand the working methodology
- One of the devices named OmniAccess generates logs in the form of natural language, So the pattern was captured using a sentiment analysis model to identify the signatures and term frequency-Inverse document frequency (TF-IDF) embedding and cosine model was used to find out a most similar article
- For other device named OmniSwitch, a lexicon-based model was created to identify the most similar article
- IMDB formula was adopted to utilize the Engineer's rating of the suggestion given by the model. So, that model can learn using feedback and recommend that article at the top if a similar issue present in the log
- Developed machine learning model to predict Estimate Time of Arrival (ETA) for analysis by utilizing log type, size and host server status using OPTUNA Auto-ML framework.
- Developed the pipeline as a python package and hosted it as API using Quart and Uvicorn ASGI framework
- Window Batch file was created to fetch the new Knowledge base article from Client CRM and dump it to MongoDB, So that package can create a model. It was scheduled for daily data retrieval and training

Tools used: Python 3.7.3 , MongoDB, Quart, Uvicorn, Window 2019 server, Anaconda.

Coeus (Sep 2020–Nov 2020)
Role: Business Analyst
Summary: Requirement gathering and interaction with all project manager to collect understand and collect dataset for the Coeus–Enterprise Insights Platform
Requirements:

- Collected and analyzed the dataset and done knowledge transfer to the development team
- Documentation for the software
- Extraction transformation and loading (ETL) process of data using python
- Development for Structured Query Language (SQL) query for Store procedure

Tools used: Python, MS–SQL server, MS–Word, MS Powerpoint, Mantis Bug tracker

Enterprise search Engine(Dec 2020–Jan 2021)
Role: Data Scientist
Summary: Developed a search engine that can search for information within the enterprise. Enterprise search systems index data and documents from a variety of different sources like CRM, knowledge base, Intranet etc.
Requirements:

- Developed python package that utilizes poppler to convert pdf to image and upload to Azure container and index it
- Developed the package to retrieve dataset from Salesforce CRM and intranet
- Used BM25L to rank the search items from different sources and give the relevance score to each result.
- Developed package for Feature snippet(FS), people also asked(POA) and Rich List(RL) for search engine
- Developed the classifier model to understand the input text feed by used and trigger FS, POA and RL API.
- Hosted entire package as Flask–CherrPy restful API and managed using PM2

Tools used: Python, sklearn, Beautiful soup, Numpy, Salesforce API, Jupyter Notebook, Azure Cognitive Services, Bing Search, pm2, VScode.

Enterprise Email classifier(Jan 2021– Ongoing)
Role: Data Scientist
Summary: Developed an API that will be utilized by Salesforce CRM to identify the queue need to route incoming ticket and assign the best agent to solve the ticket
Requirements:

- Participated in the development of application architecture and blueprint to define application component, platforms, interface and development tools
- Used TF–IDF for word embedding and LightGBM for classification
- Hyper–tunned the parameter using a Grid search
- Model is scheduled to train on new data automatically
- Hosted the API as Flask restful API

Tools used: Python, MS–Excel, Flask, AWS–Sagemaker

AWARDS AND HONORS

Postman Student expert
Date: Feb 18, 2021
Issuer: Postman

HI FLYER AWARD
Date:Jul 1, 2020
Issuer: CSS Corp

Certificate of merit
Date: Oct 1, 2017
Issuer: Bidhan Chandra Krishi Vishvavidhyalaya
description: University Gold medal

Certificate of Endowment
Date: Oct 1, 2017
Issuer: Bidhan Chandra Krishi Vishvavidhyalaya
description: Smt. Shishubala Memorial Gold Medal

CERTIFICATES

- Introduction to R Software (NPTEL)
- Scrum Foundation Professional Certificate (Certiprof)
- Machine learning (NPTEL)
- Deep learning (NPTEL)

LIST OF PUBLICATIONS

- A decision support system for the identification of critical zones in a watershed to implement land management practices (Springer)