Sumit Kumar

Bengaluru | sumit.iitkgp79@gmail.com | +91-851-389-7967

Professional Summary

Data scientist with 4+ years of experience in advanced analytics and machine learning. Proficient in developing machine learning applications and solving complex problems by extracting insights from data that help deliver value to the firm. I have experience in working on creating end to end solutions in product development using machine learning techniques.

Education

B.Tech : Electronics and Electrical Communication Engineering   *June 2013 – May 2018*
Indian Institute of Technology, Kharagpur, West Bengal CGPA – 6.12

Skill

- **Machine learning**: Linear Regression, Logistic Regression, Decision Tree, SVM,KMeans clustering, Random Forest, Gradient Boosting algorithms (GBM, XGBoost)
- **Statistical skills**: Descriptive statistics, Probability distributions, Hypothesis testing, PCA
- **Programming**: Python, SQL, Perl
- **Deep Learning and NLP**: ANN, CNN, RNN, LSTM, Transfer Learning, NLTK, Spacy, OCR Text Extraction
- **Framework**: TensorFlow, Keras, PyTorch, Flask

Relevant Experience

Data Scientist, RedKenko Health Tech Pvt. Ltd.   *Feb 2022 - Present*

- Implemented AI based OCR text extraction on **Medical lab reports** of different vendors having diverse formats and extracted key markers which will **reduce insurance claim settlement** time by 20%.
- Developed a pipeline to clean and process the data and extract key markers using **DOCAI**.
- Automated backend work using python script which reduces the hospital onboarding time on the platform by 35%.
- Implemented AI based OCR text extraction on Medical invoice bill and extracted line items which will help in **reducing OPD claim settlement** time by 25%.
- Successfully created the pipeline to flow the data into elasticsearch from a different API which will help in reducing the dependency on CRM tools like hubspot.
- Developed the AI model to classify emails which will enhance the efficiency in **assigning reimbursement emails** by 60%.

Senior AI Scientist, Data Poem - Cognitive Roi   *Nov 2020 - Jan 2022*

- Designed and developed a high performing tool **DataPoem for Media Planning** from scratch which is an offline and online planning platform for a cookie-less world. It enables the media planner to plan dynamic scenarios linked to business **KPIs.**
- Lead the team to develop Data Poem tools having Ingest & validate data , train & analyze models with other important metrics such as R2-score,**ROI**,RMSE, and deploy in production on **AWS** cloud using **ECS** and **Fargate.**
- Responsible for enhancing and maintaining the Data Poem tool, which has a pipeline to pull data from different streams and training , attribution, causality and optimisation.
- Experience on **AWS cloud** services such as **S3, EC2, ECS, DocumentDB, Fargate,Networking,Lambda function,Batch process and step function.**
- Developed Agile media planning based on accurate attribution for digital marketing using attribution algorithms.

AI Producer, Sonasoft AI India
March 2019 - Oct 2020

- Developed an **AI-based on-premise real-time system** which enables real-time prediction on stream of data.
- Responsible for development of Cognitive Intelligence Platform, where I designed and developed a stack of AI-models.
- Designed and developed a unique **Cognitive Marketing tool** for one of the leading marketing companies world-wide which enabled the company to apportion the total sales among various marketing mediums with an **accuracy of over 90%**. Further helping the company to identify the **most efficient** Marketing Mix.
- Developed a **regression AI-model** for one of the leading fertilizer companies, which would enable the company to identify key matrices involved in production of ammonia thus helping in increasing overall production.

Software Engineer, Semantics3 Pvt. Ltd.
Sept 2018 - Feb 2019

- Responsible for creating wrapper for online shopping websites. It extracted the meaningful data from the website and stored the accessible format in the database.
- Responsible for analyzing and updating the online crawler which extracted data beautifully.
- Technology used- Perl, Python, PostgreSQL, JIRA, knowledge of web development, MongoDB, XML, Elastic search.

Ideal Analytics Solutions, Summer Internship
May 2016 - July 2016

- Collected and preprocessed the data, performed feature engineering for Loan Default Prediction.
- Implemented Support Vector Machine to train the model and optimized parameters using CV error.
- Validated model using confusion matrix, precision and recall and got an accuracy of 78%**.**

Academic Projects

B.Tech Project 2:
Jan 2018 - April 2018

Facial Emotion Recognition of an Image using Deep Learning (CNN)

- Built a deep neural network model using Keras and TensorFlow to recognize emotions of images.
- Used convolution 2D and max-pooling to build a four-layer deep network.
- Achieved an accuracy of 68.04% on the test set and validated the result using Confusion matrix.

B.Tech Project 1:
July 2017 - Dec 2017

Facial Emotion Recognition of an Image using SVM Classifier

- Converted jpg images to pixel values and exported to text using MATLAB script.
- Implemented Support Vector Machine, train the model and perform facial recognition.
- Achieved an accuracy of 65.30% on the test set and validated the result using ROC curve.