MANU SIDDHARTHA

Certified Machine Learning Professional | Kaggle Kernel Expert

Mobile : (+91) 8584986762 | Email : manu.siddhartha289@gmail.com

PROFILE SUMMARY

Result oriented Machine Learning Engineer with 9+ years of experience in Banking & Finance domain. Proficiency in building end-to-end machine learning pipeline from data collection to machine learning model deployment & maintenance. Highly organized, self-motivated and have excellent communication skills.

DATA SCIENCE POC & CASE STUDY

- Capstone Project in Finance - Forecasting of Gold Rates with most comprehensive set of features using machine learning - SVR, Lasso, Ridge, Random Forest, SGD etc.
- Customer Segmentation - Analyzed a dataset of customers for a wholesale company, and applied unsupervised learning algorithms for segmenting customers into similar groups.
- Analyze Customer Feedback of Amazon Alexa - EDA & NLP
- Lending Club Case Study - Using EDA to unveil driving factors behind loan default
- Build a contextual Chatbot for restaurant search using RASA NLP
- Build a model to understand the factors on which car price vary & help Chinese company enter US market - Linear Regression & Ridge Regression.

SOCIAL LINKS

- LinkedIN : linkedin.com/in/msid-0289
- Kaggle : kaggle.com/sid321axn
- Github : github.com/sid321axn
- Personal website : www.manusid.com

ACADEMIC CREDENTIALS

Liverpool John Moores University **10/2019 - 04/2021**
MSc in Machine Learning & Artificial Intelligence

IIIT, Bangalore **10/2019 - 10/ 2020**
PG Diploma in Machine Learning & AI

Udacity **10/2018 - 02/ 2019**
Machine Learning Engineer Nanodegree

Uttar Pradesh Technical University, Lucknow **08/2006 - 06/2010**
B.Tech in Information Technology

EXPERIENCE

Indian Bank **April 2020 - Present**
Senior Manager -IT (Sr. Machine Learning Engineer)

- Worked with the team of credit analysts for building credit risk model for the bank and deployed the model for pilot basis and also submitted our research paper in international journal for the same.
- Involved in database migration of Learning Management System of the Bank from Postgres to Oracle
- Helped in optimizing Learning Management System of the Bank by extracting useful patterns of users progress and visualizing it on Power BI.
- Design end-to-end machine learning pipeline by performing data collection, data cleaning, data pre-processing, data analysis, visualization, handling outliers, performing ML model training, cross validation and hyper-parameter tuning

Allahabad Bank **May 2018 - Mar 2020**
Senior Manager -IT
(Sr. Machine Learning Engineer)

- Developed customer churn predictive model using Logistic Regression and Random Forest in AzureML and deployed as a web service on Microsoft Excel for planning department.
- Developed predictive model for targeting customers for newly launched product of the Bank based on historical marketing campaign data.
- Involved in data pre-processing, Exploratory Data Analysis, Feature Engineering, Model development, evaluation and deployment as a web service using AWS Lambda.
- Developed topic modelling based text extraction tool for Legal department of the Bank for extracting key topics of lengthy legal documents using Latent Dirichlet Allocation algorithm in python.

LANGUAGE & FRAMEWORK

- Python
- C#
- VB
- SPARK
- PIG
- HIVE
- SQL
- Javascript
- JQUERY
- AngularJS
- AJAX
- CSS3
- HTML5

PACKAGES & TOOLS

- Pandas
- Numpy
- SCIPY
- Statsmodel
- Matplotlib
- Seaborn
- Scikit-Learn
- NLTK
- Plotly
- Tensorflow
- Keras
- Pytorch
- Eli5
- PDP
- Shapley
- LIME
- Tableau

DATABASE & FILE SYSTEM

- MySQL
- SQL Server
- Postgres
- Mongodb
- HDFS
- MS Excel
- Cassandra

SOFT SKILLS

- Communicator
- Intuitive
- Motivator
- Collaborator
- Listener
- Analytical

Allahabad Bank **May 2015 - Mar 2018**

Manager -IT

(Machine Learning Engineer)

- Developed employees churn predictive model using Logistic regression and Random forest and derive driving factors which leads to employee churn that helped bank in streamline their HR Policy.
- Involved in data cleaning, data pre-processing, feature engineering task.
- Developed big data Anomaly detection machine learning model for Internet banking/ debit card transactions using advanced ensemble technique
- Involved in data modeling, model deployment, model development using CI/CD pipeline, and maintenance on different cloud environment such as AWS Lambda, AWS EC2 Instance, Heroku Cloud and Microsoft Azure cloud.
- Research new data sources and curate datasets in finance domain for statistical analysis and application of machine learning

Allahabad Bank **Sep 2011 - May 2015**

Officer-IT

(Application Development, Database Management)

- Developed Online Loan Application Filing and Tracking System (OLAFTS) for the Bank. This application helps in real-time tracking of online loan application status of the customers. Also developed reporting module with MIS. Front end - VB.NET, JQuery & Ajax Back-end SQL Server 2008.
- Developed Customer Grievance Redress module and database architecture for the bank with MIS

RESEARCH PROJECTS

1. Diagnosis of COVID-19 from Chest X-ray images using Modified Xception Net CNN (Link to Paper)

- In this project we have used modified Xception Net CNN model for extracting feature maps from Chest X-ray images.
- We have also interpret the deep CNN model using Gradient based Class Activation Maps (Grad-CAMs).
- The project implemented using python in Keras framework.

2. Detection of COVID-19 from CXR images using White Balance and CLAHE (Link to Paper)

- In this project, we have used White balance and CLAHE for image processing and depth-wise separable CNN for classifying Chest X-ray images into normal, viral pneumonia or COVID-19 induced pneumonia cases.
- The project implemented using python in Keras framework.

3. Explanatory Artificial Intelligence (XAI) in prediction of post-operative lung cancer patients (Link to paper)

- In this project, I have developed interpretable random forest model for detecting survival status of post operative lung cancer patients.
- The data used for this paper is thoracic surgery patient's data, in which data was collected retrospectively at Wroclaw Thoracic Surgery Centre in the year 2007 and 2011.

MACHINE LEARNING ALGORITHMS

- Linear Regression
- Lasso Regression
- Ridge Regression
- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine
- K Means Clustering
- Gaussian Mixture Model (GMM)
- Gradient Boosting
- Naive Bayes
- XGBoost
- CatBoost
- LightGBM
- Isolation Forest
- PCA
- T-Sne
- Meta Learning

CORE COMPETENCIES

- Data Science
- Statistics
- Data visualization
- Machine Learning
- Operations Excellence
- Statistical Analysis
- Predictive Modelling
- A/B Testing
- Model deployment
- Productivity Enhancement
- Team Management

AUTO ML FRAMEWORK

- AZURE ML
- H20 Auto ML
- Google AUTOML
- IBM Watson AutoML

OPEN PROJECTS - PROOF OF CONCEPTS

1. Malaria Parasite classification from blood sample using Deep Convolution Neural Network (View Live web app)

- In this project, I have developed a 3 layer convolution neural network to classify malaria parasites from human blood sample images.
- The dataset used for this project is available at National Library of Medicine, USA, which contains more than 27000 blood sample images of patients.
- The model is deployed on Heroku cloud using Flask REST framework and Python

2. Bengaluru House Price Prediction using Gradient Boost Regressor (View Live web app)

- In this project, I have developed machine learning based Gradient Boost Regressor model for predicting housing prices in Bengaluru
- The dataset used in this project is curated over months of prmary and secondary research by Machine Hack Community.
- The model is deployed as a web service using AWS Lambda function with docker.

3. Fake News Classification using deep learning algorithms - BERT (View Project)

- In this project, I have used different natural language processing (NLP) based machine learning and deep learning approaches including BERT to detect fake news from just seeing the news headlines.
- The dataset used in this project is the ISOT Fake News Dataset.
- Final evaluation revealed that only 1 epoch of BERT outlined all the machine learning and deep learning algorithms applied by attaining an accuracy of 98.43%

4. Real-time diagnostic web app for detection of COVID-19 from Chest X-ray images using deep CNN (View Live web app)

- In this project, I have developed a 18 layer deep depth-wise separable convolution neural network (DS-CNN) for detection of COVID-19 from chest X-ray images.
- The data for COVID-19 cases collected from Dr. Cohen's image collection git repository and for normal patients Kaggle dataset is used.
- The deep learning model is deployed as a web app on Heroku cloud using FLASK REST API and python

PERSONAL DETAILS

- Date of Birth : Feb 1989
- Languages known : Hindi, English, and Bengali
- Birthplace : Lucknow
- Address : 10/16 Kasundia 2nd Bye Lane, Shibpur, Howrah -711104, West Bengal
- Passport Details : M7818032 (Expires 09/2025)