MANU SIDDHARTHA

Senior Manager IT, Kaggle Kernel Expert

An achievement-oriented professional, targeting assignment in Data Science/Machine Learning with an organization of high repute preferably in Banking/Finance/Healthcare industry

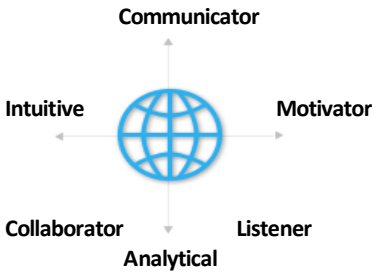
Present Location: Kolkata, India

sid321axn@gmail.com **www.manusid.com** **+91-8584986762**

Core Competencies

- **Data Science**
- **Machine Learning**
- **Operations Excellence**
- **Statistical Analysis**
- **Predictive Modelling**
- **Model Deployment**
- **Productivity Enhancements**
- **Team Management**

Soft Skills



Social Links

- **LinkedIn:** linkedin.com/in/msid-0289
- **Kaggle:** kaggle.com/sid321axn
- **Github:** github.com/sid321axn

Profile Summary

- **Result-oriented Machine Learning Practitioner** with **8.5 years'** experience across various domains such as **Banking, Finance, Healthcare and Retail industries**
- Excellent understanding of **Machine Learning Algorithms**; developed analytical methods to support novel approaches of data and information processing
- Developed AI compliant **Learning Management System, an e-learning portal** for bank employees that cut downs the cost of training by 45%
- Passionate about taking part in machine learning competitions and write Kaggle kernels for sharing knowledge and approach of solving any problem in ML domain.
- Involved in writing tutorial blogs in machine learning domain
- Interested in learning new technologies in AI domain by attending conferences, meetups and enrolling in online courses
- Proficiency in building end-to-end machine learning pipeline from building machine learning solution to deploying the solution in cloud as a web app.
- Involved in publication of research papers in AI domain by developing novel methodologies
- Possess excellent **analytical, computing interpersonal and communication skills**

Certifications

- Machine Learning Foundations: A Case Study Approach, University of Washington, 2019
- Machine Learning A-Z: Hands On Python & R, Udemy, 2018
- Machine Learning Practical: 6 Real World Applications, Udemy, 2018
- Apache Pig 101 Big Data, IBM Cognitive Class, 2018
- Accessing Hadoop Data Using Hive, IBM Cognitive Class, 2018
- Hadoop Data Access Level 1, IBM Cognitive Class, 2018

Skills

- **Programming Languages:** C#, VB.NET, Python,R, JavaScript, jQuery, HTML5, CSS
- **Database:** MS SQL Server 2008, PostgreSQL and MySQL
- **Big Data:** Spark, Hive, Apache Pig, MLLib, No SQL, MongoDB
- **Machine Learning:** Supervised Machine Learning, Unsupervised Machine Learning, Time Series
- **Deep Learning:** Multi Layer Perceptron, Convolution Neural Network, LSTM and Transfer Learning
- **Libraries:** Pandas for data wrangling, Numpy, Scikit-learn, Seaborn, Matplotlib, Tensorflow.js, Keras
- **Data Visualization:** Tableau 10, Seaborn, Matplotlib, Choropleth, Plotly and Cufflinks
- **Cloud :** AWS, Heroku and Google Cloud
- **Model Deployment:** REST API Flask, AWS Lambda (Serverless)
- **Others:** Jupyter Notebook, Google Colab, Git/GitHub

Major Projects

Project: Predictive model for increasing usage of e-learning portal of the Bank
Period: Since Jan'19
Evaluation Metrics: F1-score
F1-score Achieved: 94.58%
Role:

- Developed random forest based machine learning model for predicting which employees will complete any e-program based on their different activities in e-learning portal.
- Based on above prediction, those employees who have very less probability of completing any program in the portal are redirected to motivational page where different benefits of completing programs are listed.
- Analysed different features such as demographic, work experience and usage pattern affecting the target variable by plotting correlation plots and basic EDA
- Developed baseline logistic regression model to compare with hyper parameter tuned random forest model.
- After deploying the model we have seen around **60%** surge in number of employees who have passed an e-program.

Project: Targeting customer for bank's newly launched term deposit product using stacked machine learning solution
Period: May'18 – Dec'19
Role:

- Developed stacked ensemble of different machine learning models for predicting which customers will likely purchase newly launched bank's term deposit product based on their past transactions and basic demographic details.
- Involved in data pre-processing, normalizing and data visualization for data intuition.
- Further we have used PCA for reducing the number of dimensions and building baseline model to compare with different machine learning model.
- For increasing the accuracy of the model we have further build two level stacked ensemble solution of 10 machine learning algorithms and we have achieved 5% increase in accuracy when compared with individual machine learning algorithms
- Maximum accuracy achieved 85.93%

Project: Prediction of fraudulent Debit card and Internet banking transactions
Period: Since Jan '18
Role:

- Developed machine learning model to predict chances of fraud in debit card and internet banking transactions
- Performed Feature Engineering to add some important features which helped in increasing the model's sensitivity.
- Maximum sensitivity achieved 81.16%

Project: Prediction of churn rate of bank's customer
Period: Jun'17 – Dec'18
Role:

- Developed interpretable machine learning model using random forest to predict which customer will churn in future
- In this project, we have used different interpretation techniques to determine which factors leads customer to churn in future
- Maximum sensitivity achieved is 90.36%

Open Project: Malaria Parasite classification from blood sample using Deep Convolution Neural Network
Live at – (https://malaria-detection-app.herokuapp.com/)
Period: May'19 – June'19
Evaluation Metric: F1-score and Sensitivity
Role:

- In this project I have developed a 3 layer deep convolution neural network to classify malaria parasites from human blood sample images.
- Dataset I have used for this project is available at US National Library of Medicine (https://lhncbc.nlm.nih.gov/publication/pub9932) which contains more than 27000 blood sample images.
- Deployed the deep learning model on Heroku cloud using FLASK REST API and python.

Open Project: Bengaluru Housing Price Prediction using Gradient Boost Regressor
Live at – (https://sid321axn.github.io/bang-webui/**)**
Period: Aug'19

- In this Project I have developed machine learning based housing price predictor using Gradient Boost Regressor Algorithm and hosted on Serverless AWS Lambda function using Docker.
- The dataset is curated over months of primary & secondary research by Machine Hack (https://www.machinehack.com/) team.

Work Experience

Since Sept'11: Allahabad Bank

Growth Path: Sept'11 – May'15: Officer IT ; May'15 – May' 18: Manager-IT; Since May'18: Senior Manager IT

Role:

- Lead a team of four officers IT in building, developing and finally deploying machine learning solution.
- Managing optimization of Machine Learning and Deep Learning models
- Supervise in designing & maintenance of RDBMS and handling Big Data for Pattern Extraction and valuable insights.
- Query optimization for extracting data efficiently for creating datasets for analytical tasks.
- Performing Data Visualization using Tableau 10 and its integration with SQL Server for presenting insight reports
- Using machine learning and statistical modelling techniques to develop and evaluate algorithms to improve performance, quality, data management and accuracy
- Develop project plans and schedules for timely completion of priority projects.
- Building new models to extract more value from the customer data we collect, which will help us know more about users and merchants than anyone else in marketing or retail
- Understanding business objectives and developing models that help to achieve them, along with metrics to track their progress

Research Publications

- **Author -** Manu Siddhartha, Paramita Maity, Dr. Rajendra Nath. **Explanatory Artificial Intelligence (XAI) in the prediction of post-operative life expectancy in lung cancer patients.** *International Journal of Scientific Research***.** Vol 8, Issue 12, Dec 19; 23-28

Education

- **High school -** Happy Hours School, Lucknow affiliated to Uttar Pradesh Board - passed with 58.30% marks - in 2004
- **Intermediate** - SKD Academy Rajajipuram, Inter college, Lucknow affiliated to Uttar Pradesh Board - passed with 72% marks - in 2006
- **B.Tech. (IT)** - Saroj Institute of Technology & Management, affiliated to Gautam Buddha Technical University, Lucknow, Uttar Pradesh - passed with 76.50 % (Hons.) – in 2010
- **Machine Learning Engineer Nano Degree** - **Udacity** (Link: https://confirm.udacity.com/FDK2MPPE)
- **MS with PG diploma (Integrated Program) in Machine Learning and AI** - jointly by IIIT, Bangalore and Liverpool John Moore's University, U.K. - Pursuing - From Sept 2019 to Apr 2021

Personal Details

- **Date of Birth:** 18th February 1989
- **Languages Known:** English, Hindi and Bengali
- **Domicile:** Lucknow, Uttar Pradesh
- **Address:** 10/16, Kasundia 2nd Bye Lane, Shibpur, Howrah - 711104
- **Passport Details.** **M7818032** (Expires 03/29/25)

Other Academic Details

- Awarded with **Best Student of the year** for three consecutive years 2007, 2008 and 2009 and got scholarship during graduation.
- **TOEFL** Score – 90 in June 2019