



Sudeshna Kundu

Data Scientist

Mobile: +91-74837 44032

Mail: sudeshna.kundu@gmail.com

Key Competency Areas

- Machine Learning
- Deep Learning & Neural Networks
- Unsupervised Machine Learning
- Computer Vision & Image processing
- Natural Language Processing
- Pharma, BFSI, Education, Healthcare, E-Commerce. CPG, Travel Domain

Trainings & Certifications (Udemy)

- Artificial Intelligence: Reinforcement Learning
- Artificial Intelligence A-Z™: Learn How to Build an AI
- Unsupervised Machine Learning Hidden Markov Models
- Cluster analysis and unsupervised machine learning
- Learning path python and predictive analysis
- Complete guide to Tensorflow for Deep Learning
- Making predictions with Data and Python
- Advanced AI: Deep Reinforcement Learning
- Deep Learning and Computer Vision A-Z™
- Machine Learning A-Z™: Hands-On Python
- Data Science and Machine Learning Bootcamp
- Natural Language Processing with Deep Learning

Education

MTECH –CSE from WBUT

MBA –IT from Sikkim Manipal University

DOEACC “B”-level (Equiv to MCA) from CDAC

BSc. Physics(Hons) from Calcutta University

Higher Secondary from WBCHSE

Madhyamik from WBBSE

Professional Summary

Data Science professional with 5.10 years of working experience focused on data science, machine learning and advance analytics. Total IT experience = 10+ yrs. Key highlights are as follows:

- Deep understanding & Experience in working on both structured and unstructured data sources for various Pharma companies
- Strong expertise with **SQL, Python, TensorFlow, Keras, OpenCV, PySpark, Google Cloud APIs, Power BI**
- **Expertise in Machine Learning, Natural Language Processing, Image Processing, Predictive Modelling, Unstructured data analysis, Convolutional Neural Network, Time-series forecasting and propensity modelling, Numpy, Pandas, Scipy, Scikit-learn, NLTK, Spacy, Gensim, CoreNLP, NLP, BERT, Transformer, Audio AI/ML, Knowledge Graph, Optimization Model, ML-Ops**
- Experience with **Azure, Azure OpenAI, Cognitive Services, Azure Synapse Analytics Service, Informatica, PowerShell, ARM**, Automating Azure IaaS with PowerShell, CLI and ARM Templates, DataBricks, **Amazon AWS SageMaker**, Data Factory/Data Flow, RDBMS

Key Projects and Responsibilities

Loan Eligibility Calculation System and Fraud Detection during Loan Processing

- Developing an End-To-End system for Loan amount eligibility, loan defaulter detection/ Fraud detection during Loan Processing for a customer
- Model development, deployment and monitoring using **AWS Sagemaker** in **AWS Cloud** Architecture
- **Amazon Redshift** for saving data
- **Tools Used:** AI/ML, Python, SkLearn, Tensorflow, Keras, PyTorch
- **Client:** Leading American Bank

Organ Image Diagnostics during Hospital Endoscopic Surgery

- A medical device, endoscopic surgery probe has two **fish-eye lens** camera fitted from 2 directions, for taking images of the organ and body parts where it is being inserted into human body. The captured images are then de-noised, DE-warped and stitched to get the clarity and detection of organ medical condition, blood flow speed and directions, presence of fat globules and diagnostics of the same
- Worked as a data scientist to put together the architecture of the data flow in order to extract the attributes from the organ images through a series of process which includes image processing (de-noising, fish-eye **dewarping**, **image stitching, Homography**) and attribute extraction, on the fly video preparation and diagnostics using **MASK R-CNN, YOLOv3**
- Double Layer Model Architecture – 1st Layer for Organ detection –using YOLOv3
- 2nd Layer for Organ disease/anomaly detection – using CNN (backbone - MobileNet)
- F1-Score obtained 96.54 %
- **Amazon Redshift** for saving images and image features
- Model deployment and monitoring in **AWS Cloud** Architecture using **AWS Sagemaker**
- **Tools Used:** AI/ML, Python, OpenCV, Kornia, PyTorch, TorchVision, CUDA, SkLearn, Tensorflow, Keras
- **Client:** Largest American Multinational pharmaceutical company (worked at the client location)

Topic modeling and Correlated Topic Modeling of marketing data

- Developing an end-to-end solution to find similar documents based on simple search through investigation numbers or key phrases.
- The algorithm ranks the documents based on similarity scores from methods such as cosine distance, Jacquard distance, Word-mover distance. Prior to that, the documents vectorised through word2vec or FastText depending on the distance metric used.
- Latent Dirichlet Allocation and Hierarchical LDA, unsupervised ML model
- used for marketing data analysis and categorization, review sentiment analysis of client data
- An UI developed to expose the algorithm to the SMEs or auditor who would be requiring such a tool to fasten their audit process.
- **Tools Used:** AI/ML, Python, SQL, SkLearn, BERT
- **Client:** Largest American Multinational pharmaceutical company (worked at the client location)

Customer sentiment analysis on review of sold products

- NLP(Natural Language Processing) is used for different customer sentiment analysis on review of sold products and tagging of similar words/specs
- BERT transformer model architecture have been used
- tagging for better clarity/better categorized output
- Worked as a data scientist to put together the architecture of the data flow and models to be used to generate topics and label them on the documents
- LDA and CTM were used as the base model to create topics after data was prepared from raw free form text and cleaned
- **Tools Used:** AI/ML, Azure ML, Python, NLP, NLTK, SkLearn
- **Client:** VISTRA INC. CHINA

Employment History

Coforge Technologies

May 2022 – Till Date
Data Science
Technology Specialist
(Data Science - Team Lead)

Wipro Limited

Jan 2021 – April 2022
(Payroll company – JoulesToWatts Business Solutions Pvt LTD)
Data Scientist

Accolite Digital

July 2020 – Dec 2020
Data Scientist

Cognizant Technology Solution

Nov 2019 – April 2020
(Payroll company – Experis IT Pvt LTD)
Data Scientist, Artificial Intelligence and Analytics

Hexaware Technologies Pvt. Ltd

Apr 2019 – Oct 2019
Data Scientist

Orion India Systems Pvt Ltd

April 2016 - Mar 2019
:: (Data Scientist) Senior Software Engineer
June 2014 - Mar 2016
:: Senior Software Engineer

IBM India Pvt. Ltd.

(Payroll Company – ADECCO India Pvt LTD)
July 11 – June 14
Sr. Application Developer
Microsoft Technologies, C#, ASP.Net, MVC, Entity Framework, SQL Server)

Sales Forecasting of manufactured and marketed products

- Developed a robust hybrid Time Series forecasting model using **ARIMA** and **Time Series** decomposition, after consulting contact centre agents and analysing 4 years historical sales data, reduced variance to 5.1%
- Taking care of seasonality, weather and geographical region, diversity etc. during sales forecasting
- Used **Marketing Mix Modeling (MMM)** and **Multi-touch Attribution Modeling(MTA)** to estimate the impact of various marketing tactics
- **Tools used:** AI/ML, Python, SQL, Power BI, Informatica, GCP
- **Client:** EFG Corp. USA

Automated Chatbot For Insurance Application

- **Automated chatbot preparation for Insurance Application (with application related reply configuration)**
- **Audio processing through AI/ML**
- **Tools Used:** AI/ML, Python, Pytorch, CUDA, NLTK, Spacy, SkLearn, Tensorflow, Keras. **DeepPavlov**
- **Client:** Largest Health Insurance MNC from Indonesia and Singapore.

Customer Email classification and sending reply to emails

- NLP(Natural Language Processing) is used for different customer email classification , categorization and automated meaningful reply and taking significant action for the same.
- Sentiment analysis also had to be done for better clarity/better categorized output.
- BERT transformer model architecture have been used
- Designed the methodologies to prepare the data from the unstructured data on which several statistical test (**ANOVA**, **Apriori** rule, **PCA**) and machine learning techniques (logistic regression, random forest) were performed to test the hypotheses
- Performed sentiment analysis on incoming email data using Stanford CoreNLP package. Deep learning technique recursive neutral network was used to score the sentiment.
- Worked as a Data Scientist to put together the architecture of the data flow and models to be used to generate topics and label them on the documents
- Emails contain attached text images, pdf, client signature etc. Worked with data model in order to extract the attributes from the document through a series of process which includes OCR, language detection, language translational and attribute extraction
- **Tools Used:** AI/ML, SQL, Python, Django, NLP, NLTK, Spacy, Gensim, Rasa-Core, SkLearn, Tensorflow, Keras, GreenPlum
- **Client:** VISTRA INC. CHINA

English Handwriting recognition and digitization of Student Exam Sheet [OLICR]

- English handwritten digits and alphabets are detected from a photograph of handwritten papers snapped through the mobile camera.
- The photograph of the HandWritten paper is de-skewed and de-noised using Python, OpenCV, Pillow etc packages. HandWritten individual digits/alphabets are then cut precisely, normalized and processed, in order to tally with the initial configuration of EMNIST data sets.
- From EMNIST Training DataSet, Statistical model (CNN) is prepared in order to recognize Test DataSet and any other processed image.
- Using InfiMNIST, huge training data set with background noise is prepared.
- Processed test images are State compared and Predictive Analyzed /recognized.
- Accuracy obtained: 94.8 %
- For mobile application development of the same, Kivy is used.
- **Tools used:** AI/ML, Python, TensorFlow, Keras, OpenCV, Numpy, Pandas, Scipy, Scikit-learn, Kivy, Amazon AWS SageMaker, Matlab, SQL
- **Client:** [OLICR is the product of ORION] CBSE board schools