Siddhant Panda

Data Scientist | Gen AI | Engineer I | Samsung R & D Noida | 6x Azure Certified | Azure Certified Data and AI Engineer

Innovative Data Scientist with 3.4 Years of Data and AI Expertise, Focused on Delivering Efficiency and Quality Solutions

siddhantpanda786786@gmail.com
9599564384
Noida, India
LinkedIn
GitHub
Portfolio
Leetcode

WORK EXPERIENCE

Data Scientist, Engineer I

Samsung Research & Development Noida

07/2023 – Present *Noida*

Achievements/Tasks

- Developed an AI model achieving 93% accuracy to predict suspects in application areas based on defect data titles, problems, logs, and reproduction routes in Samsung's big data services. Integrated MLOps for constant training with new and corrected data, enhancing initial defect analysis and reducing manual efforts by 35%.
- Implemented a Gen-AI-based multi-modal AI system for generating social media captions within Samsung Gallery. Extracted data from image utilized RAG & LLM for efficient caption generation with hashtags & emoji's. Developed API for seamless on-device AI integration on mobile, winning an award for the best idea in this category.
- Developed a multi-modal AI to block age-inappropriate content on smartphones in audio, video, and text formats. Enabling real-time age detection for content classification and blocking based on user age. Implemented on device AI in mobile integration and received an award for the same.
- Developed application for on device TTS, STT, TTT feature verification comparing S24 on-device AI model with Open-Source Models. Generated detailed comparison excel reports with WERs reducing manual effort by 90% and improving accuracy by 20%.
- Developed application for multilingual relevant text extraction from images via OCR with LLM. Compared results with on-device gallery AI capture feature verification, generating Excel reports with word error rate and accuracy. Reduced manual efforts by 70% and enhanced ROI by 50%.
- Implemented video-to-audio conversion and automated S3 bucket uploads, streamlining training of videos in multiple designated categories for MLOps pipeline in Amazon Sage Maker. Generated Excel reports with detailed analysis and corrected data improved model performance by 30% and resulted in a 15% ROI enhancement.

Skills: Python (Programming Language) · Generative AI · Natural Language Processing (NLP) · Named Entity Recognition (NER) · Lang chain · RAG · Large Language Models (LLM) · Massive Multitask Language Understanding (MMLU) · General Language Understanding Evaluation (GLUE) · Transformer Models · FastAPI · Computer Vision · TensorFlow · Long Short-term Memory (LSTM) · Transfer Learning · PyTorch · Audio Processing · Text-to-Speech · Speech To Text · Multilingual Speech Processing · Pandas · Vector Databases

Data Scientist, Consultant I

Neudesic an IBM Company

04/2022 – 06/2023 *Noida*

Achievements/Tasks

- Fine-tuned and quantized the Llamma2 7B LLM model over a dataset of historical client business documents, deployed closed source LLM on TensorRT on CUDA company's GPUs and optimized inference time deployed on internal server. Developed chat-bot on stream lit & implemented Lang Chain & RAG for advanced prompts alongside LLM function calling for better user journeys for the business document
- Developed a DL model to classify the Chest-X-Ray DICOM file data into pneumonia, lung-cancer and normal CT-scan achieving 95% accuracy through hyper parameter tuning. Deployed Fast API using Azure ML & optimizing inference time.
- Developed anomaly detection in manufacturing electronic chips to detect faulty manufactured chips. Leveraging YOLOv5 and Azure Cognitive Services, to enhance the manufacturing defect prediction. Deployed ONXX, PyTorch & TFLite models for mobile, desktop & webapps deployment.
- Developed Deep Learning model for multilabel business documents classification, converting unstructured speech, audio, video, and image data to text, achieving 95% accuracy through hyper-parameter tuning. Deployed Flask API using Azure web app & optimizing accessibility.
- Designed and implemented Airflow pipelines for FTP data extraction in diverse formats. Developed custom operators for complex transformations, enhancing efficiency. Incrementally loaded data to Blob storage as Parquet files, facilitating seamless integration with Azure Synapse Analytics. Implemented SCD Type 2 queries for historical data storage in the data warehouse, ensuring high data quality and availability for business insights. Timely handling of ad hoc requests and pipeline fixes optimized operations, maximizing ROI.

SKILLS

Machine Learning, Python, SQL	AzureML,PySparkML
Gen-AI, RAG, LLM, Llamma, GPT, GANs, NER, LLMOps	Data Science, Predictive Modelling, Sckit-Learn,
NLP, Deep Learning, PyTorch ,Keras, Tensorflow, T5, TensforflowRT, Hugging Face, Transformers	Pandas, Data Engineering, Hadoop, Hive, Flume. HDFS, Spark, Airflow, Fast-API, Flask-API
PowerBI, Tableau, Data Analysis, Data Modelling Azure Data bricks, Azure Data Factory, Azure Cognitive Services, Azure Data factory	Azure Synapse Analytics, Azure Data Lake Gen2

PERSONAL PROJECTS

Lung Cancer Detection 3D Convolutional Neural Network

- Used 3D CNN to analyze low-dose CT scan data and predict the likelihood of lung cancer. Achieved 64% accuracy in classifying patients as either suffering from cancer or not

[Research Paper Published Wily Publication, M.Tech Major Project]

Image Steganography using efficient net B0, google AI

- Developed an efficient and reliable method using advanced CNN PyTorch models by Google AI, EfficientNet, to detect secret data hidden within digital images with a minimal false positive rate. Achieved an F1 score of 77.7%.

[M.Tech Research Thesis]

Fraud Credit Classification

- Goal in this classify fraudulent & non fraudulent credit card transactions. The accuracy is measured using the Area under the Precision-Recall Curve (AUPRC). Confusion matrix accuracy is not meaningful for unbalanced classification. Achieved Model Accuracy 80%

[B.Tech Minor Project]]

Image Captioning Using Deep Learning Tensor Flow

- Project aim is to generate captions for images. Dataset used a pre-trained image-model (VGG16) to generate a "thought-vector" of what the image contains, and then we trained a Recurrent Neural Network to map this "thought-vector" to a sequence of words. Achieved Model Accuracy 60%

[Internship Project]

Real Time Twitter Sentiment Analysis Using Apache Spark

- This project gives real time data (in every 5 minutes) of popular across globe or any keyword or you can use any word it will show to that tweet real time on console. Using function oriented language Scala & spark streaming library

[Internship Project]

PUBLICATIONS

The Emergence of Technology for Automated Healthcare (06/2021)

Lung Cancer Detection Using 3D CNN Based on Deep Learning

LANGUAGES

English
Full Professional Proficiency

Hindi
Elementary Proficiency

• Developed PySparkML model for credit score bracket prediction based on credit-related data, achieving 95% accuracy. Created a Flask API for ML model deployment.
• Designed Power BI dashboards to gain valuable insights from existing credit-related data for bank email marketing.
• Built Azure Data Factory pipelines using services like ADLS Gen2, Azure SQL Server, and Azure Databricks for efficient ETL processes. Leveraged Apache Spark ML to optimize data processing and model training in Azure Databricks.

Skills: Gen-AI , LLMOps · TensforflowRT · Keras · Fast-API · Langchain · RAG · Azure ML · Azure Web App · Python (Programming Language) · Natural Language Processing (NLP) · Machine Learning · Deep Learning · BERT · Transformers · TensorFlow · PyTorch · Logistic Regression · Flask · Microsoft Azure Machine Learning · Pandas · Apache Airflow · Apache Spark ML · AIOps · Extract, Transform, Load (ETL) · Pandas · SQL · Azure Databricks · Azure Data Factory · Azure Synapse. Microsoft Power BI · Extract, Transform, Load (ETL) · Predictive Analytics · Data Modeling

Data Scientist, Associate Engineer
Nagarro
03/2021 – 04/2022 *Gurugram*
Achievements/Tasks
• Boosted client fundraising campaign ROI by developing and implementing machine learning models to predict donor behavior and estimate potential donation amounts based on past profiles. Achieved 96% accuracy.
• Leveraged data analysis and data modeling expertise to accurately visualize donor behavior, optimizing fundraising efforts. Providing valuable insights for more targeted and effective campaigns, thereby maximizing ROI.

Skills: Python (Programming Language) · SQL · Pandas · Microsoft Power BI · Extract, Transform, Load (ETL) · Predictive Analytics · Microsoft Azure Machine Learning · Data Modeling

Data Scientist, Intern
TAGPAY Limited
06/2020 – 11/2020 *London (Remote)*
Achievements/Tasks
• Developed integrate game mechanics to elevate engagement and loyalty in diverse contexts. Leveraging YOLOv5 and Azure Cognitive Services, to enhance footballer performance and productivity. Deployed ONXX and TFLite models for mobile apps.
• Developed Power BI dashboards for smart watch data analysis, driving user engagement and ROI

Skills: Python (Programming Language) · Transformers · Flask · Computer Vision · Deep Learning · Microsoft Cognitive Services · Microsoft Azure Machine Learning

Big Data, Trainee
TCS iON
05/2018 – 08/2018 *Noida*
Achievements/Tasks
• Hands-on Industrial Training On Big Data Technologies: HDFS, Hive, Apache Spark, Apache Kafka, Apache Pig, Python, and Ozzie.
• Developed a real-time Twitter sentiment analysis project, updating every 5 minutes with global trending topics or specified keywords. Implemented with Scala and Spark Streaming for efficient data processing.

Skills: HDFS · Hive · Apache Spark · Apache Kafka · Apache Pig · Python (Programming Language) · Ozzie · SQL

CLOUD CERTIFICATES
Microsoft Certified: Azure Data Engineer Associate (03/2023 - 03/2025)
Microsoft Certified: Azure Data Scientist Associate (02/2023 - 02/2025)
Microsoft Certified: Azure AI Engineer Associate (02/2023 - 02/2025)
Microsoft Certified: Azure AI Fundamentals (09/2022)
Microsoft Certified: Azure Data Fundamentals (09/2022)
Microsoft Certified: Azure Fundamentals (09/2022)

COURSE CERTIFICATES
Deep Learning Specialization (Coursera)
Machine Learning Specialization (Coursera)
Hadoop Master Class (Coursera)
Machine Learning A-Z (Udemy)
Deep Learning A-Z(Udemy)
NLP Specialization (Udemy)

EDUCATION
B.Tech (Computer Science Engineering) + M.Tech Dual Degree (Artificial Intelligence & Robotics)
Gautam Buddha University
06/2016 – 06/2021
83.3%

CBSE Class 12th, Science
Kendriya Vidyalaya
06/2015 – 06/2016
82%

CBSE Class 10th, General Science
Kendriya Vidyalaya
06/2013 –06/2014
82%