

Summary:

- Success driven with 3.3yrs. of expertise in Software architecture in building algorithms and predictive models in domains such as Banking, Financial, Health Care & Insurances.
- Passionate & thriving engineer with the ability to apply ML techniques & algorithms to solve real world problems. Recognized for strength in technical skills with complete knowledge & understanding of whole SDLC, methodologies include AGILE, SCRUM.
- Implementation of AI/ML on various cloud platforms like AWS, Azure, GCP with MLOps. Technology driven executive with success in defining strategy, planning, architecting and implementing enterprise-wide initiatives.

Core Competency:

- Implementation of NLP, NLU, Generative AI, Prompt Engineering, ChatGPT using LLM Models
- Clustering & Classification, Data visualization, Predictive analysis, Statistical modeling.
- ML Algorithms, ML Model development, Data analytics, Quantitative analysis.
- Supervised/Un-Supervised Models, Reinforcement Learning, Dimensionality Reduction Algorithms, Model selection and boosting – XGBoost.
- Ability to implement libraries like sklearn, spaCy, NLTK, NumPy, TensorFlow, pytorch, keras, pandas, scipy, nltk, etc.

Skills Highlights

Machine Learning & Data Analysis	<ul style="list-style-type: none">● AI, Machine Learning – Regression, Clustering, Classification, NLP, CNN, ANN, Deep learning, Association, Reinforcement Learning, XGBoost.● Python Programming – Data science – Prep & Cleansing data, Feature Scaling● Tableau, minitab, Qlik, PySpark, Databricks, Snowflake
MLOps	<ul style="list-style-type: none">● ML model deployments & Monitoring with KubeFlow● Deployments & Monitoring in Cloud platforms (AWS, Azure, GCP)● CI/CD ML Pipelines & Auto ML implementations Cloud - AWS SageMaker, Azure ML Studio, GCP Vertex Studio
Libraries	<ul style="list-style-type: none">● SKLearn, Pandas, Scipy, Numpy, XGBoost, Apyori, TensorFlow, PyTorch, NLTK, Keras, Matplotlib, spaCy, HuggingFace, coreNLP
Containers	<ul style="list-style-type: none">● Docker, Kubernetes, Azure kubernates
Tools	<ul style="list-style-type: none">● GIT, arcmap,Terraform, YAML, flask , Jenkins, Splunk, ElasticSearch
Databases	<ul style="list-style-type: none">● Cstone DB3-internal,Oracle, JDE,Cosmos DB
Cloud Technologies	<ul style="list-style-type: none">● AWS IAM/S3/Lambda/StepFunctions/EC2/Sagemaker● Azure Machine Learning, Computer Vision, Cognitive Services, cloud, Policy Add-on/Blobs/Service Principles
Gen AI Models	<ul style="list-style-type: none">● Transformers, GANs,GPT 4,Claude 2,Texttract and Comprehend

Professional Experience
(McKinsey)-Payroll-ABCEDO Technologies Pvt Ltd
GenAI-IDP

Title: Data Scientist

Project #1

The GenAI-IDP project is an advanced AI-based solution designed to streamline and automate document processing workflows using Natural Language Processing (NLP) and Machine Learning (ML) techniques. The system integrates AWS services such as Amazon Comprehend for document classification and Amazon Textract for extracting text and structured data from various document types (e.g., invoices, receipts, bank statements). This solution is divided into three core stages:

1. Upload: Users can bulk upload files, with each document being assigned a status (successful or failed upload) and organized in S3 folders. The system tracks the status of uploaded files and displays metadata such as file name, type, timestamp, and upload status in a user interface.
2. Classification (Amazon Comprehend): Leveraging NLP, Amazon Comprehend automatically categorizes documents into predefined types (e.g., invoice, receipt, bank statement). The process is transparent to the user, with real-time feedback displayed via a processing status bar showing how many files are processed. The system adds custom metadata to each document to indicate its classification result. The classified documents are stored in S3, categorized into appropriate folders, and their status is reflected in a UI-based table for user reference.
3. Extraction (Amazon Textract): The extraction phase operates in a synchronous mode, where each document undergoes text extraction. Textract captures key-value pairs and structured data from documents. If fields in a document fall below a defined confidence threshold, they are flagged as requiring attention. The extracted data, including confidence scores, is stored in DynamoDB's "Results" table. The user can interact with the system through an editable template, where low-confidence fields are highlighted for manual review. Once reviewed, the updated documents are moved to a designated

Key Responsibilities:

- Architected and developed an AI-powered document processing pipeline using Amazon Comprehend for NLP-based classification and Amazon Textract for data extraction.
- Implemented S3-based storage solutions for organizing and managing document stages (Upload, Classification, Extraction).
- Designed a real-time file status tracking UI displaying processing status, timestamps, and document categories.
- Utilized DynamoDB for storing extracted JSON data, maintaining the results and change history with user-editable templates for low-confidence fields.
- Implemented a dynamic confidence score validation system, ensuring high-quality data extraction and allowing users to manually review and update flagged fields.
- Integrated NLP techniques to classify and extract relevant data fields automatically, improving processing speed and accuracy.
- Developed comprehensive logging and auditing features, tracking all document changes and providing an easy-to-use interface for users to monitor document history.
- Collaborated on future-proofing the system by planning for additional stages, such as data Enrichment, to further improve document processing outcomes.

Environment: Anaconda 3.6, Python, Hadoop, Hive,SQL Server. Jupyter IDE, AI/ML, NLP, XGB,LLMs,AWSSagemaker

Model output

The model will produce the probability score between 0 and 1. The scores generated indicates the likelihood of prospect to bonus game on the new US OPEN card acquired through unsolicited (Tier 2) channels. Higher the predicted score higher the changes of being a gamer. The output of this model will be used to decide whether a customer will get the welcome bonus offer for the new OPEN card or not, in case of unsolicited (Tier 2) channels.

Cligniti - Payroll-ABCEDO Technologies Pvt Ltd

Data Scientist

Project #2:

AML KYC document classification using AI/ML

Description:

- Client receives numerous amounts of KYC documents (for AML Compliance): Volume as on APR 2023 is ~15K per month.
- Tedious to be classified manually and processed by their officers. Due to which their retail customer's services in the said duration were either getting suspended temporarily which was contributing to increase in customer dis-satisfaction.
- To address the same, they have built an ML/NLP based document classification solution

Environment: Anaconda 3.6, Python, Hadoop, SQL Server. Jupyter IDE, AI/ML, NLP, Azure synapse, Azure ML Studio, Cosmos DB, RAG Architecture, GAN model, Azure blob storage, Open Cv

Responsibilities:

- Data Pre-Processing: This phase incorporated loading of data, data cleansing and enrichment; feature engineering and feature extraction, finally applying feature scaling to the whole dataset. Data labelling, converting data into either categorical/numerical value.
- Connecting to the multiple data sources by using Azure Synapse SQL queries extracting the data from data warehouses and transferring to target location
- Developed models using tensorflows, keras library and implemented hyperparameters to fine tune the model. Different models are used to handle specific problem statement.
- Implementation of Azure – Registering thing/devices, creating & registering certificates, Open Ai Search services
- Implemented Azure ML algorithm and adjusting hyper-parameters and fine tuning the model.
- Preparing training/Inference pipelines in Azure Synapse.
- Used Azure OpenAI services to achieve both the project outputs, SF to create orchestration and chaining.
- Understanding & implementation of Natural Language Processing (NLP), Natural Language Understanding (NLU) and Natural Language Generation (NLG).
- Regression model was developed based on various assumption and multi-variant variables.
- To extract text and data from the scanned documents.
- Used ML algorithms to find relationships in the extracted text.
- Deployed various applications on OpenShift with Jenkins, SonarQube.
- Demonstrated capability in machine learning techniques (e.g., classification, regression, clustering, sequence generation, forecasting, reinforcement learning, Markov modeling, or Bayesian statistics)
- Quantitatively and qualitatively evaluate new ML/AI techniques and integrate into algorithm design
- Communicate algorithm designs and performance results to cross-functional stakeholders
- Maintain concise descriptions of algorithm design specifications and performance evaluations.
- Implemented the model using Kubernetes on Azure cloud including setting up of score.py for Azure.

Birla soft- Payroll-ABCEDO Technologies Pvt Ltd

Data Scientist

Project 3

Predictive maintenance for medical devices:

Description: To optimize the maintenance and performance of medical devices in healthcare settings. The objective is to develop models and algorithms that can predict potential failures or malfunctions of medical devices, enabling proactive maintenance and reducing downtime. This approach helps healthcare providers ensure the continuous availability and reliability of medical devices, ultimately improving patient care and reducing costs.

Environment: Anaconda 3.6, Python, Hadoop, SQL Server. Jupyter IDE, AI/ML,

Responsibilities:

- Data Pre-Processing: This phase incorporated loading of data, data cleansing and enrichment; finally applying feature scaling to the whole dataset.
- Demonstrated capability in machine learning techniques (e.g., classification, regression, clustering, sequence generation, forecasting, reinforcement learning, Markov modeling, or Bayesian statistics)
- Quantitatively and qualitatively evaluate new ML/AI techniques and integrate into algorithm design
- Communicate algorithm designs and performance results to cross-functional stakeholders
- Maintain concise descriptions of algorithm design specifications and performance evaluations.
- Generated reports and graphs based on various models.

- Created charts and performed preliminary analysis & data visualization using Matplotlib.
- Implemented the model using Kubernetes on Azure cloud including setting up of score.py for Azure.

Education Certifications & Training:

BTECH Computers Science from JNTUA in 2022