Shubham Mishra

SENIOR BUSINESS ANALYST (LEAD ASSISTANT MANAGER) || EXL SERVICES || IIIT-B

Solution-driven analytics professional with four years of hands-on expertise in leveraging data-driven insights to solve complex business problems. Proficient in statistical analysis, machine learning, and data visualization techniques, with a strong foundation in programming and data manipulation. Proven track record of delivering impactful solutions through end-to-end data analysis, from data gathering and pre-processing to model development and deployment. Seeking opportunities to transform company practices into fresh, cost- effective solutions leading to more efficient operations.

EXPERIENCE

EXL SERVICES

Lead Assistant Manager – Digital Analytics Team   April 2020– PRESENT | NOIDA

Roles & Responsibilities

- Working on multiple projects that involves NLP based Leave Management Analytics, Smart Audit Systems, Repeat Call Analytics on the data that is scraped out of scanned PDF's which are converted into triples and structured format then building further predictive solutions on it.
- Developed an intelligent customer support system utilizing Language Models (LLM) and Question Answering (QA) models to enhance customer service efficiency and satisfaction.
- Created a Smart Audit System Solution in AWS by creating a structured relational dataset in RDS using Sagemaker Python Code which took the data from S3 dump. Then created multiple ensemble algorithms to solve the audits questions in an insurance flow created by our clients. Also, created multiple RestAPI's using Amazon API Gateway to solve the purpose of Deployment for front end.
- Created Repeat Call Analytics solution with an objective to solve and reduce Agent's efforts and AHT in the process of assisting customers. The descriptive analytics includes Quadrant based segmentation of High and Low performing Agents. Policy analytics Included how each policy is interlinked with each other and how one Agent can solve The purpose of repeat calls in future. Other analysis included same day repeat analysis, Lead to repeat definition and analysis, Best time to call analysis and Churn Analysis.

Birlasoft (India) Ltd.

Data Scientist – AI and Automation Team   October 2018 – April 2020 | NOIDA

Roles & Responsibilities

- Worked on Text- Analytics using natural language processing and converted tokens into NLTK data for machine processing and automation.
- Worked on Creating Complex Embeddings like Word2Vec and Glove to improve the models stability and their relationships with the context.
- Worked on Fraud Data Analytics using different Vectorizers like Count and tf-idf with the implementations of model selection techniques using parameters like Specificity, Sensitivity, Precision, Recall and F1-Score.
- Deployed customized models using Textual based algorithms like Naïve Bayes, SVM's, Logistic Regression, XGBoost, Random Forest and Decision Trees.
- Perform data exploration and data mining to create and run models.
- Worked on model improvement techniques including ridge-Lasso regularizations, Scalers, Smote and K-Fold Cross Validation.

CORE COMPETENCIES

- Python
- SQL
- Machine Learning
- Artificial Intelligence
- AWS
- Hypothesis Testing
- Segmentation
- Predictive Modeling
- Regression Modelling
- Classification Modelling
- Clustering Modelling
- Statistical Modeling
- Agile Methodology
- Strategic Planning
- Flask/Django Framing
- Alteryx

SOFT SKILLS

- Serve as an expert in translating complex data into key strategy insights and value add actions.
- Good Understanding of data architecture including data ingestion pipeline design, Data-Flow architecture and advanced data processing. Experience optimizing ETL workflows.
- Worked Exclusively on Data Cleaning and Feature Selection using standard methods like one hot encoding, PCA analysis and AutoEncoders.
- Involved in most of the phases of Software Development Life Cycle.
- Automated several hard-Coded applications in the project from inputting the Data to Exception File Generation.

Key Projects

Insurance Based Australian Project
- Led a project to automate claims processing in the insurance domain by leveraging Language Models (LLM) and Question Answering (QA) models, resulting in improved efficiency and accuracy.
- Problem Statement: Addressed the challenge of manual and time-consuming claims processing, leading to delays, errors, and increased operational costs.
- The overall architecture includes 4 modules namely dataLoader, Retriever Module, Hopping Module and Answering Module.
- The pipeline used dataLoader module to fetch the claim number for which the query is asked from SQL. Utils file was used to enhance the modularity of the code.
- Retrieval Module targets ranking the most important claim note using cross encoder approach which can be used as a context to answer the query.
- Hopping Module targets creation of sub-queries which revolves around the parent query to support and validate the outcomes from answering model for the betterment of prompts.
- Model Development and Training: Developed and trained the LLM and QA models using Python, TensorFlow, and the Hugging Face Transformers library. Conducted extensive experimentation and hyperparameter tuning to optimize model performance.
- Performance Evaluation: Achieved an accuracy of 95% in accurately identifying claim details, policy coverage, and eligibility criteria. Reduced manual effort by 70% and decreased processing time by 50% compared to manual methods.
- Hugging Face Models like T5_abs_qa, Colbert, gpt-neo-2.7B, gpt-j-6B, gpt-neox-20B, flan-alpaca-xl and alpaca-native have been used which are trained on datasets like Alpaca, MS Marco, C4 and Pile.
- Deployed the model in Domino Platform.
  Tools Used: Jupyter, Domino and Python (Transformers, Torch, TensorFlow, NumPy, Pandas, Matplotlib etc)

Pharma Based USA Project
- Created an automated solution from the huge dataset and created a model that can predict a drug to be best suited for a patient suffering from a given disease.
- Created a Causal analysis model using hypertuned Sequential Neural Network model using Glove Embedding with high stability.
- Created an adverse event analytics solution using ensemble of complex Machine learning models including SVM's, Random Forest and Logistic Regression.

EDUCATION

June 2014- June 2018
Manipal University Jaipur (Computer Science), Jaipur
CGPA: 8.2/10

April 2012 - May 2013
Bishop Johnson School, Prayagraj
HSC: 80%

April 2010- May 2011
Bishop Johnson School, Prayagraj
SSC: 83.3%

CERTIFICATIONS

May 2019 – April 2020
Machine Learning and AI, IIIT -Bangalore

Jan 2019- July 2019
Machine Learning and Deep Learning, Birlasoft

July 2017- Aug 2017
Python Programming Certification, RICE University

July 2017- Aug 2017
Python Programming – "The Python Programming Mega-Course" – Ardit Sulce

ACHIEVEMENTS

Won **Hackathon** in **Birlasoft**

Zonal Coordinator for **IIT-Roorkee**

Won **Gold Medal** for **Chess Championship Tournament**

President of **Computer Club** in **SSC**

Won **Gold Medal** in **Shikha Samvardhanam**

- Developed a Python Code for mapping the Drugs name and composition according to the Geo-location as that may vary from location to location and manufacturer to manufacturer.
- Optimized the code using Data Profiling, Coverage and PDB.
- Assisted in deploying the model in an AWS Platform.

Tools Used: Anaconda, Jupyter and Python (SKlearn, Keras, TensorFlow, NumPy, Pandas, Matplotlib etc)

Insurance Based UK Project
- Created a Fraud Detection model with high Sensitive rate which is running every month before data load and exception file generation.
- Developed a Fully Functional Business Process application with Zero tolerance to Outliers.
- Played a key role in understanding the issues measured as far as corrupt data is Concerned and reporting them to the clients on timely basis.
- Automated hard-Coded applications in the project from inputting the Data to Exception File Generation.
- Being an active part of discussions (Onsite and Offshore) to have proper understanding of the project and the requirements for the future Data Submissions played an important role.

Tools Used: Python Libraries (NumPy, Pandas, Sklearn, Matplotlib etc.), Alteryx and Excel/CSV

Celusion Technologies Pvt. Ltd. (Internship)

Data Science Intern   **June 2017 – August 2017 (0.2 Years) | THANE**

- This project in completely based on Artificial Intelligence and Machine Learning.
- Worked exclusively on Artificial Intelligence Markup language programming (AIML) and Natural Language Toolkit (NLTK).
- This includes stemming and classification of words using parts of speech as recognizer and their classification using different classifiers for confidence recognition and other rest API usages.
- Used Flask for the Deployment of the Chatbot that can deal with recommending a user what they can buy next if they have already bought an item from the inventory.
- The model was integrated in the Chatbot itself.

Tools Used: Python, Python Libraries (NumPy and Pandas), Predictive Modelling and Excel.

CONTACT DETAILS

E-Mail:
shubham.mishra0103@outlook.com

LinkedIn:
https://www.linkedin.com/in/shubham-mishra-83908a117/

GitHub:
https://github.com/shubham-mishra007

Phone Number: 8765292399