**Md Kishor Morol**

**Machine Learning Engineer / Data Scientist**

**Email: Rishitha.23g@gmail.com**

**Phone: +1 469 277 7354**

More than 8 years of experienced as machine learning engineer and data scientist in building machine learning systems for large-scale data. Skilled in a wide range of machine learning techniques, including large language models (LLMs), deep learning, NLP, and computer vision. Proficient in popular programming languages and frameworks such as Python, R, Java, C++, TensorFlow, Keras, PyTorch, and scikit-learn in order to developing end-to-end deep learning solutions for structured/unstructured data. Strong in data preprocessing, feature engineering, and model evaluation. Collaborative and effective communicator, experienced in translating business requirements into technical solutions. Proven track record of delivering impactful solutions across various industries.

**Professional Summary:**

* Proficient in building machine learning systems for large-scale data and developing end-to-end deep learning solutions for structured/unstructured data.
* Proficient in implementing deep learning algorithms such as Artificial Neural network (ANN), Convolution Neural network (CNN) and Recurrent Neural Network (RNN), tuned hyper-parameter and improved models with Python packages TensorFlow and PyTorch.
* Highly competent at wide varieties of Data Science programming languages and Big Data tools such as Python, R, SQL, Tableau, Sci-kit Learn, Hadoop, and Spark.
* Expert in DevOps and MLOPs platforms including Jira, Snowpark, Snowflake, Amazon SageMaker, MLflow.
* Expert in Data cleaning and Exploratory data analysis (EDA) and Initial Data Analysis (IDA).
* Built models with TensorFlow and acquainted with top-level frameworks such as Keras, and PyTorch.
* Processed large data sets in Python with libraries such as NumPy, SciPy and Pandas for data analysis and numerical computations to ready the data for ML algorithms.
* Skilled at Feature Engineering and Feature Selection to select subsets of relevant features to reduce the computational time without losing any data that adds value to ML.
* Hands on with various Data cleansing process like handling missing values by using feature selection such as a replacing by mean, forward or backward fill, removing entire rows or columns or values, removing outliers, normalizing, and scaling data.
* Experience with a variety of NLP methods for information extraction, topic modeling, sentimental analysis, parsing, and relationship extraction with developing, deploying, and maintaining production NLP models with scalability using NLTK, Textblob, spacy, gensim.
* Visualized data using different visualization tools R, Azure ML and Power BI.
* Developed predictive models using Random Forest, Boosted Trees, Naïve Bayes, SVM, Logistic Regression, and Neural Networks.
* Experienced in Data Integration, Validation and Data Quality controls for ELT process and Data Warehousing using MS Visual Studio, SSAS, SSIS and SSRS.
* Worked and extracted data from various database sources like Oracle, SQL Server, DB2,and migrated to Data Warehouse and Business Intelligence applications.
* Experience with relational and non-relational databases such asSQL, PostgreSQL, JSON, MongoDB, and Cassandra.
* Concrete mathematical background in Statistics, Probability, Differentiation and Integration, and Linear Algebra and Geometry.
* Acquainted with all aspects of Software Development Lifecycle (SDLC) from requirement analysis, Design, Building Coding, Testing, Deployment, and Maintenance in both Agile and Waterfall methodologies.
* Strong technical expertise at creating advanced machine learning algorithms, validation techniques, predictive data modeling, data mining algorithms, topic modeling, and sentiment analysis on the text data to provide new dimensions to the conventional thoughts on businesses.
* Interpreted and communicated the use cases, patterns, results, validity, and metrics to other teams and management in a visual and plausible way.

**Technical Skills:**

|  |  |
| --- | --- |
| **Languages** | Python, Java, C++, R, Java Script, PHP, C#, Go |
| **Packages/Libraries** | Pandas, NumPy, LangChain, Scikit-Learn, SpaCy, Seaborn, Flask, Django, LlamaIndex |
| **Machine Learning** | Supervised & Unsupervised learning, Linear Regression, Logistic Regression, Naïve bayes, Random Forest, SVM, Gradient Boosting, PCA, KNN, K-Means |
| **Deep Learning** | ANN, CNN, RNN, LSTM, GPT, BERT, ResNet |
| **DL Frameworks** | PyTorch, Tensor Flow, Keras, PySpark |
| **NLP** | NLTK, CoreNLP, Transformer, Hugging Face |
| **Big Data Frameworks**  | Apache Spark, Hadoop, Apache Flink |
| **MLOPs** | Amazon SageMaker, MLflow |
| **Data Visualization** | Tableau, Matplotlib, Seaborn, Power BI |
| **ETL/Agile Tools** | SSIS, TFS, Azure Data Factory, Data Transformation Services, Bulk Copy |
| **Database Tools** | SQL**,** PostgreSQL, T-SQL, NoSQL, Snowflake, Mongo DB, SQL Server, Hadoop/Hbase |
| **Cloud Technologies** | Azure SQL Database, Azure Data Lake, Azure Data Factory, Azure DevOps |
| **Configuration Tools** | Team Foundation Server (TFS), GitHub |
| **RDBMS** | Microsoft SQL Server, Microsoft Access |
| **Networking** | Ethernet, TCP/IP, VPN, LAN/WAN |

**Certifications & Courses:**

* Applied Data Science with Python Specialization - Offered by University of Michigan
* TensorFlow Developer Professional Certificate - Offered by DeepLearning.AI
* Natural Language Processing Specialization - Offered by DeepLearning.AI
* Deep Learning Specialization - Offered by DeepLearning.AI
* Mathematics for Machine Learning - Offered by Imperial College London

**Education:**

* American International University-Bangladesh (AIUB)

Master of Science in Computer Science (**Magna Cum Laude**)

* American International University-Bangladesh (AIUB)

Bachelor of Science in Computer Science (**Magna Cum Laude**)

**Professional Experience:**

**Client: Deloitte, NY**

**Duration: 11/2023 - Present**

**Position: Machine Learning Engineer**

**Project Description:**

I have been deeply involved in developing Generative AI projects, focusing on customizing clients' private chatbots using open-source GPT models and enhancing their interpretability with XAi tools like Captum.ai. I have leveraged LSTM in PyTorch for time series analysis and used various statistical metrics to evaluate model performance. My experience includes creating ETL mappings, performing data visualizations with Tableau, and building advanced Machine Learning models such as XGBoost, KNN, and SVM. Additionally, I have handled NLP tasks using techniques like Word2Vec and tf-idf, managed data transformations in Hive and Hadoop, and developed distributed TensorFlow environments. I have also managed AWS EC2 instances, utilized JIRA for project tracking, and collaborated with business partners for continuous improvement.

**Roles & Responsibilities:**

* Working on **Generative AI** projects for developing **Large Language Models (LLMs**) using open source Generative Pre-trained Transformer (GPT) to customize clients private chat bots.
* Developing **Explainable AI (XAi)** models with Interpretability tool **Captum.ai** so that clients can understand how their model performs.
* Used **Long-Short Term Memory** (**LSTM**) for analyzing time series data in **PyTorch.**
* Proficient at statistical metrics like **F-Score, AUC/ROC, Confusion Matrix** and **RMSE** to evaluate different model **performance.**
* Generated **ETL** mappings, sessions and workflows based on business user requirements to stack data from source files, **RDBM** Stables to **target tables.**
* Performed **data visualization** using various libraries and designed dashboards with **Tableau** generated complex reports, including charts, summaries, and graphs to interpret the findings to the team and stakeholders.
* Built advanced **Machine Learning** classification models like **XGBoost, KNN, SVM, Regression** and clustering algorithms **Hierarchical Clustering** and **DBSCAN.**
* Performed **NLP** by using techniques like **Word2Vec, Text Blob, Bagof Words, tf-idf, Doc2Vec.**
* Performed data analysis by using **Hive** to retrieve the data from **Hadoop** cluster, **SQL** to retrieve data from **Oracle** database.
* Handled importing data from various data sources, performed transformations using **Hive, Map Reduce**, and load data into **HDFS.**
* Analyzed data to identify glitches and cleaning it to reduce the distortions.
* Applied multiple **Machine Learning** and **Data Mining** techniques to improve the quality of product ads and personalized recommendations.
* Developed **NLP** with **Deep Learning** algorithms for analyzing text improving over their existing dictionary-based approaches.
* Created distributed environment of **TensorFlow** across multiple devices (**CPU’s** and **GPU’s**) and run them in parallel.
* Managed **AWSEC2** instances using Autoscaling groups and used ticketing tools like **JIRA** to monitor work.
* Created various types of data visualizations using **Tableau**, **PowerBI**, and other libraries like Matplotlib, Seaborn, ggplot2.
* Collaborated with internal business partners to identify needs, recommended improvements.

**Environment:** Python, Scikit-Learn, Git, NLP, Keras, TensorFlow, Oracle DB, SQL Server, PySpark, NumPy, Pandas, Tableau, Matplotlib, Seaborn

**Client: Classy Girl Beauty Supply Inc, Fort Lauderdale, FL**

**Duration: 06/2023 – 04/2024**

**Position: Data Scientist**

**Project Description:**

I developed predictive analytics models using NLTK in Python to forecast demand for beauty products, optimize inventory levels, and reduce stockouts. I created dashboards and reports with Tableau and Power BI to analyze customer behavior, product trends, and operational efficiency. Using LSTM in RNNs, I performed sentiment analysis on customer feedback. I experimented with classification algorithms like Logistic Regression, SVM, and Random Forest, and evaluated their performance for customer discount optimization. I also applied various Machine Learning and Data Mining techniques to enhance product ads and personalized recommendations, and developed NLP models with Deep Learning algorithms to analyze text. Additionally, I employed statistical tests, performed data manipulation, and created visualizations using Tableau and other tools.

**Roles & Responsibilities:**

* Developed **predictive analytics** models using **NLTK** on Python to predict demand for beauty products, optimize inventory levels and reduce stockouts.
* Developing **dashboards and reports** for large datasets to extract insights of customer behavior, product trends, and operational efficiency using **Tableau, Power BI**.
* Performed **Sentimental analysis** on the product feedback of the customers to determine the emotional tone behind the series of words and gain the express of the attitudes and emotions by Long-Short Term Memory (**LSTM)** cells in **Recurrent Neural Networks (RNN).**
* Experimented with multiple classification algorithms, such as **Logistic Regression**, **Support Vector Machine**, **Random Forest**, **Ada boost** and **Gradient boosting** using **Python, Scikit-Learn** and evaluated the performance on customer **discount optimization.**
* Analyzed data to identify glitches and cleaning it to reduce the distortions.
* Applied multiple **Machine Learning** and **Data Mining** techniques to improve the quality of product ads and personalized recommendations.
* Developed **NLP** with **Deep Learning** algorithms for analyzing text improving over their existing dictionary-based approaches.
* Collaborated in building of **high-performance low latency system** to manage high velocity data streams.
* Employed **statistical tests** such as hypothesis testing, t-test, confidence intervals, error measurements.
* Performed various **data manipulation** techniques in statistical analysis like missing data imputation, indexing, merging, and sampling.
* Developed **Hive** UDF's to bring all the customers emails into a structured format.
* Created different **charts** such as Heatmaps, Bar charts, Line charts.
* Worked in creating different visualizations in **Tableau** using Bar charts, Line charts, Pie charts, Maps, Scatter Plot charts, and Table reports.

**Environment:** Python, NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Tableau, Power BI, NLP, Neural Networks, Oracle Database, SQL, AWS, Git, MS Excel

**Client: BJIT GROUP, Dhaka, Bangladesh**

**Duration: 04/2021 - 04/2023**

**Position: Machine Learning Engineer**

**Project Description:**

I performed Exploratory Data Analysis (EDA) to maximize dataset insights, detect outliers, and extract important variables. I implemented dimensionality reduction techniques like PCA and t-SNE, and developed clustering algorithms for market segmentation. Using Pandas, NumPy, and Scikit-learn in Python, I built machine learning models, including Random Forest and SVM, to predict customer churn and interface, ensuring accuracy with cross-validation and ensemble methods. I handled text data preprocessing and feature engineering, applied NLP techniques, and implemented LSTM networks with TensorFlow. I communicated results effectively using Tableau dashboards and collaborated with data science, marketing, and operations teams for informed decision-making.

**Roles & Responsibilities:**

* Performed Exploratory Data analysis **(EDA)** to maximize insight in to the dataset, detect the outliners and extract important variables by graphically and numerically.
* Implemented algorithms such as **Principal Component Analysis (PCA)** and **t-Stochastics Neighborhood Embedding (t-SNE)** for dimensionality reduction and normalize the large datasets. developed various **Clustering algorithms** for market segmentation to analyze the customer behavior patterns.
* Used **Pandas, NumPy, seaborn, SciPy, Matplotlib, Seaborn, Scikit-learn**, **NLTK in Python** at various stages for developing machine learning model.
* Implemented machine learning algorithms, **Random Forest** and **Support vector machines** to predict the Customer churn and Customer interface.
* Used **cross-validation techniques** to avoid the overfitting of the model to make sure the predictions are accurate and measured the performance using **Confusion matrix** and **Classification report.**
* Improved accuracy using Ensemble methods of the model with different **Bagging** and **Boosting** methods.
* Involved in various pre-processing phases of text data like **Tokenizing**, **Stemming, Lemmatization** and converting the raw text data to structured data.
* Performed feature engineering, performed NLP by using some techniques like **Word2Vec, BOW** (Bag of Words) **, tf-idf, Word2Vec**, **if-idf**,
* Implemented **LSTM layer network** of moderate depth to gain the information in the sequence with help of **TensorFlow.**
* Created **distributed environment** of **TensorFlow** across **multiple devices** (CPUs and GPUs) and run them in parallel.
* Used Tableau to convey the results by using dashboards to communicate with team members and with other data science teams, marketing, and engineering teams.
* Communicated the results with operations team for taking best decisions.

**Environment:** NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Tableau, SQL, Linux, Git, Random Forests, SVM, t-SNE, PCA, Tensor Flow, K-Means, Natural Language Tool Kit, LSTM, RNN

**Client: Dynamic Solution Innovators (DSi), Dhaka, Bangladesh**

**Duration: 08/2018–07/2021**

**Position: Software Engineer (AI)**

**Project Description:**

I developed efficient chatbots using ML algorithms like Naive Bayes, Decision Trees, SVM, RNN, and LSTM, along with NLP techniques such as bag-of-words and lemmatization. I applied machine learning and statistical modeling, including decision trees, text analytics, and social network analysis, using Scikit-learn in Python. I built and tested ensemble models like Random Forest, XGBoost, and AdaBoost to enhance accuracy and stability. Using LSTM, I predicted failure probabilities over time, and employed anomaly detection techniques with SVM, k-means, and k-NN. Working in an agile environment, I managed project timelines, communicated with clients, and utilized tools like Pandas, NumPy, Keras, and TensorFlow for predictions.

**Roles & Responsibilities:**

* Worked on creating efficient chatbots using popular ML algorithm such as Naive Bayes, Decision Trees, Support Vector Machines, Recurrent Neural Networks (RNN), Markov Chains, Long Short-Term Memory (LSTM) and Natural Language Processing (NLP), bag-of-words model and lemmatization.
* Application of various machine learning algorithms and statistical modeling like decision trees, text analytics, natural language processing (NLP), supervised and unsupervised, regression models, social network analysis, neural networks, deep learning, SVM, clustering to identify Volume using scikit-learn package in python.
* Built and tested different Ensemble Models such as Bootstrap aggregating, Bagged Decision Trees and Random Forest, Gradient boosting, XGBoost and AdaBoost to improve accuracy, reduce variance and bias, and improve stability of a model.
* Used LSTM to predict probability of failure at different time intervals compensating for independent variables reflecting states of wear.
* Worked in the agile environment to implement agile management ideals such as sprint planning, daily standups, managing project timelines, and communicate with clients to ensure project progress satisfactorily.
* Detected Anomaly using data science concept of Support Vector Machine (SVM), k-means Clustering, K-nearest neighbor.
* Pandas, NumPy, Keras, Scikit-learn, SciPy, TensorFlow was utilized to predict categories dependent on location, time and some different highlights of Linear regression, Logistic regression, Decision Tress.

**Environment:** NumPy, Pandas, Matplotlib, Seaborn, SciPy, Scikit-Learn, Tableau, Natural Language Tool Kit, Keras, TensorFlow.

**Client: DeepChain Labs LLC, Dhaka, Bangladesh**

**Duration: 05/2016–09/2018**

**Position: AI Engineer**

**Project Description:**

I performed data extraction and manipulation on large relational datasets using SQL and Python, collaborating with senior AI teams to integrate a machine learning layer into the final product. I trained data with classification models like Decision Trees, Random Forest, and Logistic Regression to classify quartiles and predict scores. I developed methods to consolidate and analyze unstructured data, applying various ML algorithms for actionable insights, and conducted sentiment analysis on social media using NLP techniques like word2vec and RNNs. I ensured model accuracy and low false positive rates for text classification and sentiment analysis.

**Roles & Responsibilities:**

* Performed data extraction and manipulation over large relational datasets using SQL, Python, and other analytical tools.
* Closely worked with senior Artificial Intelligence Team to create and build a Machine learning layer in the final Product.
* Trained Data with Different Classification Models such as Decision Trees, Random Forest, Linear & Logistic Regression, KNN models to classify quartiles & predict scores
* Developed and designed program methods, processes and systems to consolidate and analyze the unstructured data from multiple sources to generate actionable insights and solutions. datasets, graphics and executive level reporting for mass distribution. Application of various machine learning algorithms and statistical Modeling like decision trees, text analytics, natural language processing (NLP), supervised and unsupervised, regression models, social network analysis, neural networks, deep learning, SVM, clustering to identify Volume using Scikit- learn package in python.
* Worked on sentiment analysis on social media that includes both clustering and classification of messages.
* Worked on word2vec and deep learning for NLP algorithms such as RNN to get best results for sentiment analysis. Also Determining Information extraction from customer review and feedback. Identify and assess available machine learning and statistical analysis libraries (including regressors, classifiers, statistical tests, and clustering algorithms). Ensure that the model has low False Positive Rate and Text classification and sentiment analysis for unstructured and semi-structured data.

**Environment:** NLTK, SQL, Python, NumPy, Pandas, Keras, PyTorch

**Honors & Award:**

* Vice Chancellor’s Gold Medal for best undergraduate Project & Thesis
* Magna Cum Laude for Graduate CGPA
* Magna Cum Laude for Graduate CGPA
* Merit-Based Academic Scholarship for maintaining academic excellence throughout the undergraduate and postgraduate period.

**Selected Publications:**

1. MA Ali, **Md Kishor Morol**, MF Mridha, N Fahad, MS Al Huda, N Ahmed, "Exploring a Novel Machine Learning Approach for Evaluating Parkinson’s Disease, Duration, and Vitamin D Level", 2023
2. **Md Kishor Morol**, S Rokon, Md Jamil, "Food Recipe Recommendation Based on Ingredients Detection Using Deep Learning", 2022
3. MF Mridha, **Md Kishor Morol**, MA Ali, MSH Shovon, "convoher2: A deep neural network for multi-stage classification of her2 breast cancer", 2022
4. A Islam, **Md Kishor Morol**, SY Shin, "A Federated Learning-Based Blockchain-Assisted Anomaly Detection Scheme to Prevent Road Accidents in Internet of Vehicles", 2022
5. **Md Kishor Morol**, SS Das, S Mahmood, "Data Security and privacy in cloud computing platforms: A comprehensive review.", 2022