Data Analyst/Data Scientist- Machine learning

**Email id: pramitmaharjn@gmail.com**

**Contact: 2815440844**

**PROFESSIONAL SUMMARY**

* Around **5+ years of experience** in AI, Machine Learning and DataScience.
* Experience with large datasets of **Structured** and **Unstructured data**, Data Visualization, Data Acquisition and modelling.
* Proficient with predictive modeling, **NLP/ Deep Learning / Computer vision/ Probabilistic Graphical Models / Inferential statistics/ Graph/Data Validation**.
* Hands on experience in **data mining algorithms** and approach.
* Experience in developing different Statistical **Machine Learning**, **Text Analytics**, **Data Mining solutions** to various business generating and problems data visualizations
* Worked on **PANDAS,** **Numpy**, **Scikit**-**Learn**, **NLTK**, **Tensorflow**, **keras**.
* Experience in **Statistics**, **Regression**- **Linear**, **Logistic**, **Poisson**, **Binomial, Neural Network**
* Proficient in Machine Learning techniques (**Decision Trees, Linear, Logistics, Random Forest, Naïve Bayes, SVM, Bayesian, XG Boost, K-Nearest Neighbors**) and Statistical Modeling in Forecasting/ Predictive Analytics, Segmentation methodologies, Regression based models, Hypothesis testing, Factor analysis/ PCA, Ensembles.
* Highly skilled in using visualization tools like seaborn, PowerBI for creating dashboards.
* Strong (OOP)programming expertise in **Python, JAVA,** and strong **in Database SQL**.
* Good experience in error and exception handling.
* Worked and extracted data from various database sources like **mysql, ms Sql Server, Postgre sql**.
* Experience with file systems**, databases, SQL, and data movement (ETL).**
* Proficient in **Python**, experience building, and product ionizing end-to-end systems
* Knowledge of **Django, Rest, Springboot** and other web development tools
* Captured trends, seasonality patterns through time series models such as ARIMA, used lag variables and sliding window techniques, feature engineered variables through iterations.
* Used Git hub, Bit bucket, Jira integration tools
* Hands on experience with **pyspark, MLlib**
* Worked on Agile Environment

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| Statistics/ML | **Exploratory Data Analysis:** Univariate/Multivariate Outlier detection, Missing value imputation, Histograms/Density estimation  **Supervised Learning:** Linear/Logistic Regression, Lasso, Ridge, Elastic Nets, Decision Trees, Ensemble Methods, Random Forests, Support Vector Machines, Gradient Boosting, LDA, XGB, Deep Neural Networks, Bayesian Learning  **Unsupervised Learning:** Principal Component Analysis, K-means,  **Time Series:** ARIMA, Exponential smoothing |
| Machine Learning /  Deep Learning | **Python:** Pandas, NumPy, scikit-learn, SciPy, Statistical modeling,  **Spark:** MLlib, TensorFlow  **SQL** |
| Databases/ETL/Query | MySQL, Ms SQL Server, Postgres, SQLAlchemy and Big Query |
| Visualization | Power BI, Seaborn, Matplotlib |
| Big Data Tool | Apache Spark, Storm |

**PROFESSIONAL EXPERIENCE**

**People’s United Financial, Bridgeport, CT Feb 2018 - Present**

**Role: Data Scientist/Machine Learning**

**Description:** Working on Risk Modelling Fraud detection focussed on Credit card fraud detection modelling algorithms with Python and Spark programming in the Azure cloud environment. Have a good experience working with deep learning neural networks used for the modelling.

**Responsibilities:**

* Implemented Machine Learning, Deep Learning and Neural Networks algorithms using **TensorFlow Deep Learning Framework** and designed Prediction Model using Data Mining Techniques with help of **Python**, and Libraries like NumPy, SciPy, Matplotlib, Pandas, **scikit-learn**
* Used **scikit-learn** in modelling various classification, regression and clustering **algorithms** including [support vector machines](https://en.wikipedia.org/wiki/Support_vector_machine), [random forests](https://en.wikipedia.org/wiki/Random_forests), [gradient boosting](https://en.wikipedia.org/wiki/Gradient_boosting), [k-means](https://en.wikipedia.org/wiki/K-means_clustering), **ANN, SOM**.
* Involved in defining the source to target **data mappings**, **business rules**, and **data definitions**.
* Defining the list codes and code conversions between the source systems and the data mart using Reference Data Management (**RDM**).
* Conducted data cleaning, data preparation, and outlier detection
* Finding insights from millions of customer chat and calls records
* Gathering requirements from business
* Reviewing business requirements and analyzing data sources
* Developed predictive models for sales and Finance teams using various ML and DL algorithms
* Presented to the higher management the discovered trends and analysis, forecast data, recommendations, model results and risks identified
* Created and presented executive dashboards to show the patterns & trends in the data using PowerBI Desktop, query mining.
* Was able to identify emerging issues using the models
* Involved in developing and testing the **SQL** Scripts for report development, Power BI reports, query editor, Dashboards and handled the performance issues effectively

**Client: Walmart, Columbus, OH. June 2016 - July 2017**

**Role: Data Scientist/Machine Learning**

Developed an algorithm that accurately predicts the demand of products and overall sales prediction among multiple classes based on the historical sales data available on multiple products. Further, Maintaining the right stock of products whose demand is high while avoiding the loss of maintaining unnecessary products.

**Responsibilities:**

* Gathering business requirement from client and approach formulation and design methodology.
* Performed Data Cleaning, features scaling, features engineering using python
* Replacement of missing data and perform a proper EDA to understand the time series data
* Captured trends, seasonality patterns through time series models such as ARIMA, used lag variables and sliding window techniques, feature engineered variables through iterations.
* Involved working on different databases such as Oracle and SQL sever of different platforms etc.
* Involved working in Data science using Python 3.x on different data transformation and validation techniques like Dimensionality reduction using Principal Component Analysis (PCA) and A/B testing, Factor Analysis, testing and validation using ROC plot, K- fold cross validation, statistical significance testing.
* Check the existence of trend and seasonality in the data
* Used Python 2.x/3.x to develop many other machine learning algorithms such as ARIMA, HYBRID that help in decision making using Keras, TensorFlow and Scikit-learn.
* Improvise the model until achieving best accuracy
* Generated visualizations using Power BI(Dashboards) and Seaborn.
* Helped in Web App development
* Involved in NLP project using NLTK, Beautiful Soup, google Dialog flow

**Vertiv, Columbus, Ohio April 2015 – May 2016**

**Role: Data Analyst/Data Scientist**

**Description**: Worked on Stock Prices Prediction using LSTM Deep Neural Networks and involved in worked on Big data processing with Hadoop and creating Data Pipelines.

**Responsibilities:**

* Assisting business by being able to deliver a machine learning project from beginning to end, aggregating and exploring data, building and validating predictive models and deploying completed models to deliver business impacts to the organization
* Performed Data Cleaning, features scaling, features engineering using pandas and NumPy packages in python and build models.
* Created impact documents specifying changes introduced as part of the program and lead the business process team.
* Worked on different data formats such as CSV, JSON, XML and performed Machine Learning algorithms in Python and Deep Learning techniques such as RNN and LSTM
* Work with big data consultants to analyze, extract, normalize and label relevant data using Statistical modeling techniques like **Logistic regression, decision trees, Support vector machine, Random forest, Naive Bayes** and **neural networks**
* Review business data for trends, patterns or casual analysis to assist in identifying model drift and retraining models
* Developed MapReduce/Spark Python modules for machine learning & predictive analytics in Hadoop’s (Storm) on AWS
* Rapid model creation in Python using pandas, NumPy, Scikit-Learn, Keras
* Extracting the source data from Oracle tables, **MS SQL Server**, sequential files and **excel sheets**.
* Performed **SQL query** for data analysis and integration **Environment:** Python, HTML5, OLTP, Random Forests, OLAP, HDFS, JSON, MySQL, NumPy, Matplotlib, Spark

**Dator Technologies Inc., Kathmandu, Nepal Jan 2013 - Dec 2013**

**Role: Data Analyst**

**Description**: Performed data analysis and data profiling using complex SQL on various sources systems including Oracle and MS SQL.

**Responsibilities:**

* Involved in defining the source to target data mappings, business rules, data definitions
* Involved in defining the business/transformation rules applied for sales and service data.
* Worked with project team representatives to ensure that logical and physical **ER**/**Studio** data models were developed in line with corporate standards and guidelines.
* Define the list codes and code conversions between the source systems and the data mart.
* Responsible for defining the key identifiers for each mapping/interface
* Perform coding migration, database change management & data management through the various stages of the development life cycle
* Perform and manage daily database maintenance, monitoring and performance tuning tasks/jobs
* Implementation of **Metadata Repository, Maintaining Data Quality, Data Clean-up** procedures, Transformations**, Data Standards, Data Governance program, Scripts, Stored Procedures**, **triggers** and execution of test plans
* Remain knowledgeable in all areas of business operations to identify systems needs and requirements.

**EDUCATION**

Ms Finance from University of Bridgeport (2019)