**Pratik Lamsal**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Professional Experience:**

* Over 6 years of extensive experience in Machine Learning solutions to various business problems and generating data visualizations using Python.
* Experience as a Professional Qualified Data Scientist in Data Science and Analytics including Machine Learning, Data Mining, and Statistical Analysis
* Used Pandas, NumPy, Scikit-learn in Python for developing various machine learning models.
* Hands on experience in implementing Naive Bayes, Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, neural networks, Principle Component Analysis and good knowledge on Recommender Systems.
* Implemented deep learning models and numerical Computation with the help of data flow graphs using Tensor Flow Machine Learning.
* Worked with numerous data visualization tools in python like matplotlib, seaborn, ggplot, pygal
* Experience in designing visualizations using Tableau software and publishing and presenting dashboards, Storyline on web and desktop platforms.
* Used the version control tools like Git 2.X
* Worked and extracted data from various database sources like Oracle, SQL Server, DB2, and Teradata.
* Good knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, Secondary Name Node, MapReduce concepts, and ecosystems including Hive and Pig.
* Strong experience in Software Development Life Cycle (SDLC) including Requirements Analysis, Design Specification and Testing as per Cycle in both Waterfall and Agile methodologies.
* Proficient knowledge in statistics, mathematics, machine learning, recommendation algorithms and analytics with excellent understanding of business operations and analytics tools for effective analysis of data.
* Highly self-motivated, enthusiastic, and result-driven with the ability to effectively communicate with all levels of the organization including senior management and executives.
* Guide the development teams to break down large and complex user story into simplified versions for execution.

**TOOLS AND TECHNOLOGIES:**

|  |  |
| --- | --- |
| **Machine Learning** | **Regression analysis, Decision Tree, Random Forests,**  **Naïve Bayes, SVM, Neural Network, K-Means Clustering and KNN.** |
| **Programming Languages** | **Python (2 &3), SQL, Java, C/C++** |
| **Frameworks and Toolkits** | **NumPy and Pandas.** |
| **Statistical Methods** | **Descriptive statistics, Hypothesis Testing, ANOVA, Chi-square testing,**  **F-tests, Confidence Intervals, Bayes Law, Dimensionality Reduction,**  **ROC curve.** |
| **Databases** | **Oracle 10g/9i/8i, MS SQL Server 2008, MS Access, MySQL 5.x,** |
| **Operating Systems** | **Windows NT/2000/XP/7,8&10, Windows Server 2008.** |

**Education**

**Masters in Data science|**

**Harrisburg university|**

**2018**

**Regions Bank, Birmingham, Al March 2017-Present**

**Machine Learning Engineer/Data Scientist**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Description:** Analyzed, interpreted, & communicated financial data. Performed data mining on large datasets (structured and unstructured), build models and do analysis using advance modeling techniques. Ascertained and executed plans to build, implement and maintain predictive models using available tools and programming languages. Monitored and provided feedback on model performance and recalibrate model as necessary.

**Responsibilities:**

* Worked with several R packages including knit, dplyr, Spark, R, Causal Infer, Space-Time.
* Coded R functions to interface with Caffe Deep Learning Framework.
* Used Pandas, NumPy, Seaborn, SciPy, Matplotlib, Scikit-learn, and NLTK in Python for developing various machine learning algorithms.
* Worked on different data formats such as JSON, XMLand performed machine learning algorithms in Python.
* Setup storage and data analysis tools in Amazon Web Servicescloud computing infrastructure.
* Implemented end-to-end systems for Data Analytics, Data Automationand integrated with custom visualization tools using R, Mahout, Hadoop and MongoDB.
* Worked as Data Architects and IT Architects to understand the movement of data and its storage and ERStudio9.7.
* Utilized Spark, Scala, Hadoop, HBase, Cassandra, MongoDB, Kafka, Spark Streaming, MLLib, Python, a broad variety of machine learning methods including classifications, regressions, dimensionally reduction etc. and Utilized the engine to increase user lifetime by 45% and triple user conversations for target categories.
* Used Spark Data frames, Spark-SQL, Spark MLLib extensively and developing and designing POC's using Scala, Spark SQL and MLlib libraries.
* **Environment:** MySQL, Cassandra, Netezza, Linux, SQL, Hadoop, SPSS, Spark, Keras, TensorFlow, AWS, Pig, Hive, Windows 7, MS Excel, Tableau, VBA, R, Shiny, Python, Google Cloud, IBM Jazz, JIRA.

**Verizon Wireless - TX, USA April 2014 - Jan 2017**

**Python developer and Automation**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Description:**

Worked with programming, data mining, large scale data acquisition, transformation and cleaning of structured and unsecured data. Created predictive / prescriptive analytical models to improve the employee Experience in the Digital Workplace. Worked with data scientist to build machine learning model for natural language processing.

**Roles and Responsibilities:**

* Requirement gathering, Data analysis of different types of data.
* Creating excel files using different Python libraries Openpyxl, XLWT, Pandas.
* Implemented Porter Stemmer (Natural Language Tool Kit) and NLP bag of words model
* (CountVectorizer) to prepare the data.
* Using Pandas data frames for working with data and manipulating data.
* Automation of creation of excel sheets using Pandas based on different requirements and pulling of data from other excel files as input.
* Pandas, Numpy, Openpyxl, XLRD, XLWT.
* Extraction of data from different APIs using RESTful API services.
* Used Python AWS SDK for dynamic creation of EC2 instance and installed the application on AWS EC2 instances.
* Debug existing automation code and test to confirm functionality within AWS/EC2.
* Evaluated business requirements and prepared detailed specifications that follow project guidelines

Required to develop written programs.

* Implemented python code for retrieving the Social Media data.
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.
* Creating MySQL tables, loading with data and writing SQL queries which will be used by python programs.
* Worked on SQL joins to aggregate social media data according to the requirement.
* Created tables in MySQL which are used as staging table for the data loaded from the inbound feeds.
* Worked on different file formats like Text files and ORC file.
* Implemented Docker build tool for scheduling, deploying Python jobs.

**Environment:** Python 2.7, Pandas, Numpy, RegEx , AWS, RESTful services, HTML, JSON, XML, MySQL, all Excel related python libraries (XLRD,XLWT,Openpyxl), OS module.

**Client: NBT Bank, Norwich, NY Jan 2013- Feb 2014**

**Role: Data Scientist**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Description:** NBTBank is an independent community bank based in Norwich, NY. The project involved working with end users and analysts to provide analysis of credit card customer base on demographical basis and forecasting risk. The permissions to grant a credit card to an individual were based on the analysis done. The information of the person applying for the credit card is gathered and processed for further approval or rejection.

**Responsibilities:**

* Involved in Analysis & Marketing Team to make business decisions
* Involved with key departments to analyze areas and discuss the primary model requirements for the project
* Documented methodology, data reports and machine learning model results and communicated with the Project Team / Manager to share the knowledge
* Performed competitor and customer analysis, risk and pricing analysis and forecasted results for credit card holders on demographical basis
* Used Machine Learning algorithms and Natural Language Processing(NLP) for response modeling and fraud detection efforts for Credit cards
* Developed needs-based segmentation that aided management in gaining a deeper understanding of consumer behavior. These segments assisted management in development and marketing of credit cards.
* Performed machine learning to estimate the probability of a new customer being classified as a good or bad customer.
* Design, develop and produce reports that connect quantitative data to insights that drive and change business.
* Supported client by developing Machine Learning Algorithms on big data using PySpark to analyze transaction fraud, Cluster Analysis etc.
* Perform ad hoc custom analysis as needed using SQL and R.
* Designed and published visually rich and intuitively interactive Tableau workbooks and dashboards for executive decision making.
* Maintain and enhance data model with changes and furnish with definitions, notes, reference values and check lists.

**Environment:** DB2, Teradata, Hadoop, Spark, SPSS, Pig, Hive, XML, Excel, MS Access, VBA, Tableau, Python, R, Linux, SharePoint, HP Quality Center.