**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.** |

**Regions Bank, Birmingham, Al Oct 2018-Presemt**

**Machine Learning Engineer/Data Scientist**

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

**Description:** Regions Financial Corporation is a bank and financial services company headquartered in the Regions Center in Birmingham, Alabama. The company provides retail and commercial banking, trust, securities brokerage, mortgage and insurance products and services.

**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.
* Installed and used **Caffe NLP Framework.**
* Worked on different data formats such as **JSON**, **XML** and performed machine learning algorithms in Python.
* Setup storage and data analysis tools in **Amazon Web Services** cloud computing infrastructure.
* Implemented end-to-end systems for **Data Analytics**, **Data Automation** and 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.

**ARNIC – Cedar Rapids, IA Mar 2014-Aug 2018**

**Machine Learning/Data Scientist**

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

**Description:**

Aeronautical Radio, Incorporated (ARINC), is a major provider of transport communications and systems engineering solutions. ARINC has more than 3,200 employees at over 120 locations worldwide. The goal was creating customer profiling models and customer value analysis. Also improving customer services by automating some of the tasks using machine learning, pattern analytics and exploratory analysis.

**Responsibilities:**

* Developed Python modules, machine learning & predictive analytics for day to day business activities.
* Perform Exploratory analysis, hypothesis testing, cluster analysis, correlation, ANOVA, ROC Curve and build models in Supervised and Unsupervised Machine Learning algorithms, Text Analytics & Time Series Forecasting
* Implemented Porter Stemmer (Natural Language Tool Kit) and **NLP** bag of words model (**CountVectorizer**) to prepare the data.
* Implemented number of customer clustering models and these clusters are plotted visually using Tableau legends for the higher management.
* Developed Natural Language Processing to automate the classification of customer incident queries into levels of classes to improve the customer services.
* Implemented a machine learning model for **customer** **sentiment** **pattern** to better assess the heartbeat of the customer trend.
* Conducting studies, rapid plots and using advanced data mining and statistical modeling techniques to build a solution that optimizes the quality and performance of data.
* Demonstrated experience in design and implementation of Statistical models, Predictive models, enterprise data model, metadata solution and data lifecycle management in both RDBMS, Big Data environments.
* Developed Simple to midlevel Map Reduce Jobs using hive and Pig and developed multiple MapReduce jobs in python for data cleaning and preprocessing.
* Analyzing large data sets apply machine learning techniques and develop predictive models, statistical models and developing and enhancing statistical models by leveraging best-in-class modeling techniques.
* Worked with several outlier algorithms like **Z-score, PCA, LMS**, and **DBSCAN** to better process the data for higher accuracy.
* Worked with parameter tuning and model evaluation techniques Confusion Matrix, Cross validation, AUC-ROC etc. Customer Profiling models using **K-means** and **K-means++** clustering algorithms to enable targeted marketing.
* Developed the model with ~1.4million data points and used the **elbow method** to find the optimal value of K using Sum of Squared error as the error measure.
* Designed and implemented a probabilistic churn prediction model with ~80k customer data to predict the probability of customer churn out using Logistic Regression in Python. Client utilized the results in the business to finalize the list of customers to provide a discount.

**Client: NBT Bank, Norwirch, 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.