**Anmol Shrestha – Data Scientist / Machine Learning Expert**

**PROFESSIONAL SUMMARY**

* 7+ years of experience in IT in which more than 5 years dedicated to Data Science, Machine Learning, Data mining of large structured and unstructured data, Predictive modeling, Data Validation and Visualization.
* Experience in building solutions for enterprises, context-awareness, pervasive computing, and/or application of machine learning.
* Extensive experience in building data models by using machine learning techniques like Clustering Analysis, Market Basket Analysis, Association Rules, Naïve Bayes, Recommendation System, Dimension Reduction, Principal Component Analysis (PCA), Decision Tree and Neural Networks.
* Hands-on experience in managing all project stages including business development, estimation, requirements determination, gap analysis, business process reengineering, issue resolution, configuration, training, go-live assistance, vendor management and post implementation support.
* Experience in managing project as Technical Lead of Data Science project and handled team of 15.
* Proficient in Text Analytics, developing different Statistical Machine Learning, Data Mining solutions to various business problems and generating data visualizations using R and Python.
* Experience in utilizing analytical applications like R and Python programming language to identify trends and relationships between different pieces of data, draw appropriate conclusions and translate analytical findings into risk management and marketing strategies that drive value.
* Good knowledge on statistical analysis techniques like Confidence Interval, Hypothesis testing, ANOVA, Conjoint analysis, sentiment analysis and semantic analysis.
* Exposure to LSTM networks and Deep Reinforcement Learning. Team player with ability to communicate findings in data and policies with non-technical stake holders.
* Good Hands-on experience working with large datasets and Deep Learning class using TensorFlow and Apache Spark. Worked with Spark Core, Spark ML, Spark Streaming and Spark SQL.
* Adapt and deep understanding of Statistical modeling, Multivariate Analysis, model testing, problem analysis model comparison, optimization and validation.
* Hands on experience in working with UML, Class Diagram and Sequence Diagram.
* Good exposure on using various tools like Eclipse IDE, SQL Developer in implementing the applications.
* Ability to work in tight schedules and efficient in meeting deadlines.
* Excellent Analytical, Communication and Interpersonal skills.
* Experienced in developing applications using Agile Scrum methodology.
* Adapt and adhere to industry standards while working with multi-cultural teams.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| Languages | Python, R, MATLAB, Core java, SQL, Bash |
| Machine Learning | Linear Regression, Multiple Linear Regression, Polynomial Regression, Logistic Regression, Naïve Bayes Classifier, Clustering, Association, Decision Tree, Random Forest, Support Vector Machines (SVM), Neural Networks, Nearest Neighbor |
| Language Skills | Data Extraction, Data Pre-processing, Data Visualization, Statistical Analysis, Predictive Analysis, GGplot, Web Scraping, Functional & Object-Oriented Programming |
| Tools & Framework | Scala, Spark, Cassandra, Scikit-learn, Shogun, Accord Framework, DL4J, MLlib, Apache Singa, ND4J, Django, Flask, Spring |
| Modelling/ Case Tools | Microsoft Visio 6.0, Sybase Power Designer16.5, CA Erwin Data Modeler 7.1/4 |
| OLAP/ BI/ ETL Tools | Crystal Enterprise 10(CMC), Business Objects 6.1, Integration Services (SSIS), Analysis Services (MS OLAP, SSAS), Reporting Services (SSRS), Oracle 9i OLAP, Performance Point Server (PPS), DTS. MDX, Crystal Report 10, |
| Web Technologies / Services | HTML5, XHTML, CSS3, XML, AJAX, SOAP, RESTful, UDDI, WSDL, JAX-RPC, JAX-RS JAX-WS, JAX-B, |
| Web/App Servers | WebSphere, Apache, Tomcat, IIS ADMIN, IIS 6.5/ 7.5 |
| Database | Oracle 11g/10g/9i, Teradata, Hadoop, Cloudera Oryx, SQL-Server, MySQL server, TOAD, H2O, Mahout |
| IDEs | PyCharm, Spyder, Jupyter Notebook, Eclipse |
| Version Control | Git, GitHub, GitLab, SVN, TFS, Microsoft Visual SourceSafe |
| OS & Environment | Windows 10/8/7/XP, Sun Solaris, UNIX, Linux, VirtualBox |

**PROFESSIONAL EXPERIENCE**

**Capital One Bank, Plano, TX Jan 2017 – Current**

**Role: Data Scientist / Machine Learning Expert**

**Description:** The purpose of the project was to store terabytes of log information generated by website and processing it to get insights.

**Responsibilities:**

* Involved in various phases of analytics project lifecycle including data extraction/preparation, design and implementation of scalable machine learning analysis and solutions, and documentation of results.
* Prepared customer data by analyzing peak and off-peak time periods for the purposes of designing rates.
* Applied various machine learning algorithms and statistical modeling like decision trees, regression

models, clustering, SVM to identify volume using scikit-learn package in Python.

* Conducted cluster analysis to generate segmented profiles of customers.
* Partner with technical and non-technical, infrastructure and platform teams to configure, tune tools, automate tasks and guide the evolution of internal big data ecosystem; serve as a bridge between data scientists and infrastructure/platform teams.
* Worked on Text Analytics and Naive Bayes creating word clouds and retrieving data from the platforms.
* Support various business partners on a wide range of analytics projects from ad-hoc requests to large-scale cross-functional engagements.
* Handled analytical problems with an appropriate blend of statistical/mathematical rigor with practical business intuition.
* Pro-actively analyzed data to uncover insights that increase business value and impact.
* Prepared Data Visualization reports for the management using Python or R and Tableau.
* Worked on different data formats such as JSON, XML to perform machine learning algorithms in Python.
* Identified root causes of problems and facilitated the implementation of cost-effective solutions with all levels of management.
* Performed K-means clustering, Regression and Decision Trees in Python and also worked with Naïve Bayes and skilled in Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, Principle Component Analysis.
* Hold a point-of-view on the strengths and limitations of statistical models and analyses in various business
* contexts and can evaluate and effectively communicate the uncertainty in the results.
* Implemented end-to-end systems for Data Analytics, Data Automation and Integration.
* Pro-actively analyzed in multiple ways to evaluate approaches and compare results.

**Environment:** Anaconda Python, R, Java, PL/SQL, SQL connector, SparkML, UML, SOA, JSON, Log4J, Oracle 11g, Linux/Unix, Bash Scripting

**Blue Cross Blue Shield, Chicago, IL Sep 2014 – Dec 2016**

**Role: Data Scientist / Machine Learning Engineer**

**Description:** This product is developed with intents to improve risk management practices and install process that is more efficient related to legal data and documentation.

**Responsibilities:**

* Responsible for data identification, collection, exploration & cleaning for modeling, participated in model development.
* Integrated with other departments to extract data from complex sources for reports and analysis.
* Responsible for design and development of advanced R/Python programs to prepare transform and

harmonize data sets in preparation for modeling.

* Used Python and Spark to implement different machine learning algorithms including Generalized Linear Model, SVM, Random Forest, Boosting and Neural Network
* Created various types of data visualizations using Python.
* Implement various statistical techniques to manipulate data (missing data imputation, principle component
* analysis and sampling) and build predictive models.
* Developed MapReduce/Spark Python modules for machine learning & predictive analytics in Hadoop.
* Implemented a Python-based distributed random forest via Python streaming.
* Handled ad-hoc requests for business needs and provided high quality and accurate results.
* Experience in analyzing plans and descriptions of findings for research protocols regulatory reports and healthcare manuscripts.
* Built standard reports for company presentation, provided ad-hoc query and analysis support, and created

requirements document by interacting with customers.

* Used GIT for version control with Data Engineer team and Data Scientists colleagues.
* Worked closely with end-user clients to understand their reporting needs and provided necessary solution.

**Environment:** R, Python, UNIX Scripting, SAS, Cassandra, Java, Hadoop, MapReduce, HDFS, Oracle, Eclipse, Clustering, Regressions Analysis, Oracle, GIT

**MIO Partners, New York, NY Nov 2012 – Aug 2014**

**Role: Machine Learning Data Engineer**

**Responsibilities:**

* Participated in data acquisition with Data Engineer team to extract historical and real-time data.
* Created reports and dashboards to explain and communicate data insights, significant features, models scores and performance of new recommendation system to both technical and business teams.
* Worked closely with project manager to develop work plan for the projects and keep the manager aware of any issues.
* Provided analytical network support to improve quality and standard work results
* Provided estimates for project development and implementation. Work with management to meet those estimates.
* Develop framework, metrics and reporting to ensure progress can be measured, evaluated and continually improved.
* Designed lean proofs of concepts (POC) to answer targeted business questions.
* Made several machine learning pipelines and applied decision tree analysis, using Python.
* Implemented data processing and statistical techniques such as sampling techniques, estimation, hypothesis testing, time series, correlation and regression analysis using Python.
* Used Natural Language Processing tools / libraries like NLTK, Spacy, LDA genism for analyzing text scrapped from the web.
* Utilized tools and packages, such as Jupyter Notebooks, to generate compelling, interactive visualizations from diverse data sources.

**Environment:** Python, NLTK, Spcay, PL/SQL, Decision Tree, Machine Learning Algorithms (Linear Regression, Logistic Regression, Naïve Bayes Classifier, Support Vector Machines)

**AMNIL Technologies, Kathmandu, Nepal Jan 2011 – Oct 2012**

**Role: Python Developer**

**Responsibilities:**

* Involved in building database Model, APIs and Views utilizing Python, to build an interactive web-based solution.
* Optimized use of data types and object-oriented concepts for making complex algorithms of networks.
* Optimized database performance and use of resources, including selecting optimal physical implementations of databases.
* Automated performance calculation tasks by writing Python scripts using Numpy and SQL alchemy.
* Involved in profiling the application using Python profilers to check overall performance of a module.
* Performed data pre-processing like cleaning, integration, transformation and reduction tasks using Pandas.
* Responsible for debugging and troubleshooting the web application.
* Developed test scripts and performed Unit Testing on the applications using Django Tests.
* Supported user groups by handling target-related software issues/service requests, identifying/fixing bugs.
* Managed code versioning with GitHub and deployment to staging and production servers.
* Worked in team environment and participated in an on-call rotation schedule for production servers.
* Performed Extraction, Transformation and Loading (ETL) of data from various external sources.
* Co-ordinated with the Sr. Developer to implement any new feature that may boost the performance.

**Environment:** Microsoft Windows Server 2008 Enterprise Edition, Microsoft Windows 2003, Microsoft SQL Server 2005, 2008, 20082 and Reporting Services.

**REFERENCES**

Available upon request