**Govardhana P**

+1 (347) 471-0536 govi.0815@gmail.com

**Professional Experience:**

* Over 6 years of experience as a Professional Qualified **Data Scientist/Data Analyst** in Data Science and Analytics including Machine Learning, Data Mining, and Statistical Analysis
* Extensive experience in **Machine Learning** solutions to various business problems and generating data visualizations using Python.
* 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.

Willing to relocate: Anywhere

**PROFESSIONAL EXPERIENCE:**

**L.L. Bean - Freeport, ME Apr 2017 - Present**

**Role: Data Scientist**

**Description:** L.L. Bean is a global company sourcing its products from the US and across the globe. The company involves manufacturing and selling products such as Maine Hunting Shoe, L.L. Bean Boot, Boat and Totes, dog beds, leather goods and backpacks. The job involves creating statistical machine learning models for fraud detection, implementing automated customer scoring systems, sentiment analysis etc.

**Responsibilities:**

* Involved in all phases of data acquisition, data collection, data cleaning, model development, model validation, and visualization to deliver **data science solutions**.
* Created classification models to recognize web request with product association in order to classify the orders and scoring the products for analytics which improved the online sales percentage by 13%.
* Used **Pandas, NumPy, Scikit-learn** in Python for developing various machine learning models such Random forest and step-wise regression.
* Worked on **NLTK** library in python for doing sentiment analysis on customer product reviews and other third party websites using web scrapping.
* Used **cross-validation** to test the models with different batches of data to optimize the models and prevent overfitting.
* Implemented and developed **fraud detection** model by implementing a Feed Forward Multilayer Perceptron which is a type of ANN.
* Worked with ANN (Artificial Neural Networks) and BBN (Bayesian Belief Networks).
* Used pruning algorithms to cut away the connections and perceptrons to significantly improve the performance of back-propagation algorithm.
* Hands on experience in Dimensionality Reduction, Model selection and Model boosting methods using **Principal Component Analysis** (PCA), **K-Fold Cross Validation** and **Gradient Tree Boosting.**
* Implemented a structured learning method that is based on search and scoring method.
* Customer segmentation based on their behavior or specific characteristics like age, region, income, geographical location and applying Clustering algorithms to group the customers based on their similar behavior patterns.
* Created and maintained reports to display the status and performance of deployed model and algorithm with **Tableau**.
* Worked with numerous data visualization tools in python like **matplotlib, seaborn, ggplot, pygal**.

**ARINC – Ceder Rapids, IA Aug 2016 – Mar 2017**

**Role: 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.
* Implemented dimensionality reduction using **Principal Component Analysis and k-fold** cross validation as part of Model Improvement.
* Implemented Pearson's Correlation and Maximum Variance techniques to find the key predictors for the Regression models.
* Worked with numerous data visualization tools in python like **matplotlib, seaborn, ggplot, pygal**.

**Infosys Ltd, India**

**Client: Blue Cross Blue Shield of MN Jun 2014 –May 2016**

**Role: Data Scientist**

**Description:**

Blue Cross Blue Shield Association (BCBSA) is a federation of 36 separate United States health insurance organizations and companies, providing health insurance in the United States to more than 106 million people. In this project we built various machine learning model to help the organization in determining the good customer base, future insurance claims based on current customer information and other analytics. Used machine learning algorithms which turn images of text into editable documents and extract semantic meaning from those documents to return accurate results.

**Responsibilities:**

* Worked with **OCR** libraries extensively in order to extract the semantic data from the insurance related documents to help the preprocessing step for claims.
* Created and implemented python modules for filtering images using image processing libraries like **pillow, scikit-image, OpenCV, scipy, pycario and simpleITK** to extract text and background.
* Worked with **pytesseract** and several scikit-image sub packages like **regionprops, label, clear\_border, threshold\_otsu and hog others.**
* Evaluated models using Cross validation, Log loss function used to measure the performance and used **ROC** curves and **AUC** for feature selection.
* Implemented a deep learning based OCR method using **tensor** **flow**, **TFANN** module.
* Implemented clustering mechanism to group various classes of data such as Policy numbers, Hospital ID, Doctor ID, etc.
* Implemented **NLP** (Natural Language Processing) based classification to categorize various claims.
* Developed a machine learning model to match the claims with the supporting documents in order to decrease the manual intervention.
* Created various types of data visualizations using **Matplotlib**, **Seaborn** and **Tableau**.
* Used **Tableau** to convey the results by using dashboards to communicate with team members and with other data science teams, marketing and engineering teams.
* Created visualization maps for a diabetic insurance claims using a **heatmap** in order to find correlations with help of **seaborn** visualization.

**Infosys Ltd, India**

**Client: AMWAY Apr 2013 - May 2014**

**Role: Data Analyst**

**Description:**

The Project main objective is to create a data warehouse for all the market places of AMWAY. AMWAY has total of 56 market places each place has its own local IT team to represent and each place has its own data source. The aim is to design and implement a project which will receive the data from various data sources of these 56 market places and validate it to load it into the data warehouse. Along with these operations, the transformation of data, creating and delivering business reports to the business users is also a major part in the project.

**Responsibilities:**

* Designed, Developed and Deployed application with my team in timely manner.
* Created **Unix** **Scripts** for receiving and extracting the data files which is been sent by various market places around the world.
* Designed and implemented business rules from the scratch by a number of discussions with client in daily manner.
* Adhered to design standards and client guidelines.
* Designed a **data** **flow** **model** for each market, starting from external tables to the very end of summary tables.
* Collaborating with Product Managers and Software Engineers to provide guidelines on solid application design.
* Designed **PostgreSQL** modules for each market to validate and transform data while loading it into the database.
* Worked with **pgAdmin** administrator to design storage allocation and load balancing.
* Created and implemented daily running jobs based on the timelines with a parallel partner Cognizant.
* Tuned every data retrieval and updating process for better performance.
* Proposed some design ideas to eliminate redundant problems like cyclic dependency, campaign dependency and some data flow errors.
* Created an interface for the reporting team to access the data from the warehouse to create business reports for the users.
* Worked with reporting team to automate some of the business reports for the higher management review.
* Trained support team on every data flow error, logical discrepancies and reporting data issues.
* Created and implemented some of the new complex market places to the application as standalone resource.
* Tuned most complex markets to **gain** **28%** of performance boost on selected markets.
* Re-designed some of the markets to **cut down the critical issues by 83%.** Received a best performer award for this particular achievement.

**Infosys Ltd, India**

**Client: Westpac Banking Corporation May 2012 - Mar 2013**

**Role: Data Analyst**

**Description:**

Westpac is an Australian bank and financial-services provider. Westpac has 14 million customers, and employs almost 40,000 people. Job involves collecting data from various data sources and pump it through informatica workflows to store it into the data warehouse. This project also involves data correction, business logic implementation using PL/SQL and other scripting languages like Shell scripting

**Responsibilities:**

* Involved in Data mapping specifications to create and execute detailed system test plans. The data mapping specifies what data will be extracted from an internal data warehouse, transformed and sent to an external entity.
* Worked closely with stakeholders to understand, define, document business questions needed.
* Review system/application requirements (functional specifications), test results and metrics for quality and completeness.
* Designed and Developed **Oracle PL/SQL** Procedures and **UNIX Shell Scripts** for Data Import/Export and Data Conversions.
* Analyzed the source data coming from different sources (**SQL Server, Oracle and also from flat files like Access and Excel**) and working with business users and developers to develop the Model.
* Have Used **Informatica** Data Quality as **ETL** tool to transform the data from various sources and bring them into one common format and load them into target database for the analysis purpose from Data Warehouse.
* Executed SQL queries to validate actual test results and match expected results as per financial rules.
* Responsible for maintaining the integrity of the SQL database and reporting any issues to the database architect.
* Design and model the reporting data warehouse considering current and future reporting requirement
* Involved in the daily maintenance of the database that involved monitoring the daily run of the scripts as well as troubleshooting in the event of any errors in the entire process.