**Mario Torres**

Data Scientist

[Mario.Torres81@outlook.com](mailto:Mario.Torres81@outlook.com)

Tampa, FL

# **Professional Summary**

* Over 7+ years of experience in data science, data analysis, machine learning, predictive model building, data visualization, and statistical analysis.
* Experience in building intuitive products and experiences for millions, while working alongside an excellent, cross-functional team across Engineering, Product, and Design.
* Expert in transforming business requirements into analytical models and designing algorithms.
* Proficient in developing data mining and reporting solutions that scale across a massive volume of structured and unstructured data to improve business performance in every aspect.
* Experience working with machine learning supervised algorithms – **Linear Regression, Logistic Regression, Linear Discriminant Analysis (LDA), Decision Tree, Random Forest, Support Vector Machines (SVM), Naïve Bayes, K – Nearest Neighbor.**
* Experience working with machine learning un-supervised algorithms – **Hierarchical clustering, K-means clustering, Probability Clustering, Density-Based Clustering (DBSCAN).**
* Experience using **Dimensionality Reduction Techniques like Principal component analysis (PCA), Linear Discriminant Analysis (LDA) Independent component analysis, Random component analysis,** and **t – SNE.**
* Experienced in developing deep-learning models like **Artificial Neural Networks – Multilayer Perceptron’s (MLPs), Convolutional Neural Networks (CNN),** **Recurrent Neural Networks (RNN)** using **TensorFlow** for pattern recognition,prediction analysis, machine translation, social network filtering image & video recognition.
* Worked on **Long Short-Term Memory (LSTM)** using **Keras** forautospeech recognition and anomaly detection.
* Experience in using **Artificial Neural Networks** for recommendation systems.
* Proficient with **Natural Language Processing (NLP)** for **Interactive Voice Response (IVR), Language Translation** and **Word Processors** for grammatical accuracy of texts**.**
* Experienced working on **Natural Language Processing (NLP) techniques like Word2Vec, BOW (Bag of Words), Tf–idf, Avg-Word2Vec, If-idf, Weighted Word2Vec.**
* Used **Sentiment Analysis** to determine the emotional tone behind the series of words and gain the express of the attitudes to analyze the market of a product, customer service, fraudulent activities.
* Expert level mathematical knowledge on **Linear Algebra, Probability, Statistics, Stochastic Theory, Information Theory**, and **logarithms**.
* Experience with **Word Embeddings, Topic Modeling using Latent Dirichlet Allocation (LDA), Sentiment Analysis, Text Classification, Semantic Analysis and Parts of Speech Tagging.**
* Strong experience with **Python** and its libraries **Pandas, NumPy, Sci-Kit learn, Seaborn, Matplotlib** and **R** for algorithm development, data manipulation, analysis, and visualization.
* Proficient in writing complex **SQL** queries like stored procedures, triggers, joints, and subqueries to access and manipulate database systems like **MySQL, PostgreSQL.**
* Experience in using **Tableau** for data visualization and designing dashboards for publishing and presenting storyline on web and desktop platforms.
* Proficient in the entire project life cycle and actively involved in all the phases including data acquisition, cleaning, engineering, feature scaling, feature engineering, statistical modeling, and visualization.
* Experienced on working different file formats like **JSON, CSV, XML** in **Anaconda Navigator, Jupyter Notebook, Visual Studio code,** and **Spyder.**
* Experience in designing and deploying **AWS** Solutions using **EC2, S3, EBS, Elastic Load Balancer (ELB), auto scaling groups, optimizing volumes** and **EC2 instances.**
* Experience in creating multiple **VPC** instances, configuring and networking of **Virtual Private Cloud (VPC).**
* Created **Lambda** function to start/stop **AWS** resources based on scheduled events triggered by **CloudWatch** events to save the cost on non-prod environments.
* Knowledge of development and deployment of **machine learning algorithms & AI system** to drive real-time forecasting, personalization, and recommendation using **Amazon SageMaker** and **Spark.**

# **Tools & Technologies**

|  |  |
| --- | --- |
| Languages | Python, R, SQL |
| Packages | Pandas, NumPy, SciPy, Scikit-learn, Matplotlib, Seaborn, NLTK, Tensor Flow, Keras |
| Database | MySQL, PostgreSQL, DynamoDB, Aurora |
| Cloud Services | Amazon Web Services (AWS) |
| Mathematical Skills | Statistics, Linear Algebra, Probability |
| Machine Learning Algorithms | Linear Regression, Logistic Regression, Linear Discriminant Analysis (LDA), Decision Trees, Random Forests with Adaboost and Gradient Descent Boosting, Support Vector Machines (SVM), Naïve Bayes, K – Nearest Neighbor, Hierarchical clustering, K-means clustering, Density based clustering (DBSCAN). |
| Machine Learning Techniques | Principal Component Analysis, Single Value Decomposition, Data Standardization Techniques, L1 and L2 regularization, RMS prop, Hyperparameter tuning, KL Divergence, Resampling Techniques like SMOTE, Cluster Centroid Methods, Ensemble Methods, Feature selection and Feature Engineering, Cross Validation Methods(K-fold), Bleu Score. |
| Deep Learning | Convolution Neural Network, Recurrent Neural Network, LSTMS, GRU, Autoencoders, Generative Adversarial Neural Networks, Policy based and Value based Boltzmann Machines. |

# **Professional Experience**

## Verizon – Tampa, FL Feb 2018 – Present

### Role – Data Scientist

Responsibilities:

* Involved in the development of algorithms for fraud detection, customer churn prevention, lifetime value prediction, product development and prediction analysis based on company requirements and goals.
* Participated in all phases of data mining – data collection, data cleaning, data manipulation, developing models, validation, visualization and performed gap analysis.
* Performed **Exploratory Data Analysis (EDA)** to categorize and organize data based on caller information like identification number, date, time, type of Service (voice call, SMS, etc.), duration, network access point identifiers.
* Worked on user profiling to extract information like customer behavior & new attributes for anomaly detection.
* Worked on **K-means** clustering to find the groups of data within the number of groups represented by the variable to find the feature similarities for behavioral segmentation.
* Built **time – series** model for complex pattern recognition of financial time series data and forecast of returns.
* Performed exponential smoothening on multivariate time series data for short-term forecasts.
* Performed Sentiment analysis using **Natural Language Processing (NLP)** model on the email feedback and reviews of the customers to determine the emotional tone behind the series of words and gain the express of the attitudes and emotions by Long-Short Term Memory (**LSTM)** cells in **Recurrent Neural Networks (RNN).**
* Extracted texts data related to fraud cases from different telecommunication companies to train the machine learning algorithm with **Word Segmentation, Part of speech tagging, select keywords,** and **frequency value.**
* Trained algorithms are used to test for the word distribution, correlation value by passing it through the **NLP** (speech recognition and conversion) algorithm to detect for the possibility of fraudulent activities.
* Developed and applied machine learning algorithms linear regression, logistic regression, multiple regression, mean-variance, dummy variable, Poisson distribution, Naive Bayes, fitting function.
* Developed clustering algorithms Hierarchical, K-means with help of Sci-kit learn and SciPy to group customers and made data driven decision on promotional offers & price strategies that reduced customer churn significantly.
* Used python to build machine learning algorithms by importing Sci-kit learn, SciPy, NumPy, Pandas modules to analyze the terabytes of data to find the customer lifetime value prediction and the possibility of customer churn.
* Worked on Support Vector Machines (SVM), clustering models, Principle Component Analysis (PCA) with different structured & unstructured datasets for dimensionality reduction & analyze the accuracy of the models.
* Worked on Naïve Bayes and Random Forests, to find the possible hidden patterns for forecast predictions.
* Validated models using cross-validation and loss function to measure model performance. Created Confusion Matrix, ROC and CAP curves.
* Addressed overfitting and underfitting by tuning hyperparameters using **L1** and **L2 Regularization.**
* Involved in the development and deployment of machine learning algorithms & AI system to drive real-time forecasting, personalization, and recommendation using **Amazon SageMaker** and **Spark.**
* Visualized results in python using Matplotlib, Seaborn libraries of Scikit-learn and used Tableau to create the interactive dashboards to present results for team members, management and relevant departments in company.

## Macy’s – Duluth, GA Jan 2017 – Dec 2017

Role – Data Scientist

Responsibilities:

* Analyzed the customers purchase data and product trends to recommend the types of products/services to customers based on their behavior tracked through the customer accounts, purchase history and location.
* Acquired years of sales data from relevant and novel data sources using **SQL** queries to understand the customer purchase patterns in different quarters.
* Suggested product inventory levels based on various considerations to avoid delays on product deliveries.
* Assisted marketing team to devise the business strategy to target customers with discount coupons, deals and offers to improve customer purchases by identifying distinct patterns in which customers respond to offers.
* Clustered customer actions using **K-means** **Clustering** and **Hierarchical Clustering** and segmented them into different groups which helped the marketing team to further analyze behavioral patterns of customers.
* Used Multi-Linear Regression algorithm and created the **Customer Lifetime Value (CLV)** from the data recorded through applications for a period of at least three months.
* Identified high & low value segments to help the employer to understand customers & improve customer service.
* Developed a machine learning system that predicted purchase probability through offers based on customer's real-time location data and past purchase behavior. These predictions are being used for mobile coupon pushes.
* Developed a model that collects data across thousands of locations to optimize product placement and advertising to catch the eye of shoppers that fit the right profile and selection of products to be removed entirely to reduce clutter and make it easier to find in-demand items automatically.
* Participated in developing deep learning model for the process of comparing items against each other, tracked their performance in various situations & made suggestions to support key business decisions.
* Built one class **Support Vector Machine (SVM)** and **Principal Component Analysis (PCA**) algorithms for anomaly detection of fraud and other errors that signal dishonest behaviors.
* Forecasted sales and improved accuracy - **(MAPE and RMSE)** by 30% by implementing advanced forecasting algorithms that were effective in detecting seasonality and trends in the patterns.
* In addition to incorporating exogenous covariates, increased accuracy of the machine learning algorithms helped business plan better with respect to budgeting and sales and operations planning.
* Measured the price elasticity for products that experienced price cuts and promotions using regression methods. Based on the elasticity, made selective & cautious price cuts for certain licensing categories.
* Used python to build machine learning algorithms by importing Sci-kit learn, SciPy, NumPy, Pandas modules to analyze the terabytes of data.
* Evaluated the performance of different models using **F-score, AUC/ROC, Confusion Matrix and RMSE/MSE** and used Matplotlib extensively to generate human-readable data visualizations.

## JPMorgan Chase – Chicago, IL Jan 2016 to Dec 2016

### Role – Data Scientist

Responsibilities:

* Analyzed the data using various machine learning algorithms whether to extend/not credit limit to an existing applicant and to approve/not new credit line to a new applicant will likely result in profit or loss based on various circumstances like credit history, utilization rate, income, age, location, hard enquiries & number of deliquesces.
* Extracted terabytes of structured and unstructured data by using **SQL** queries and performed **data mining** tasks including handling missing data, data wrangling, feature scaling, outlier analysis in python by importing pandas.
* Conducted data investigation, discovery & mapping tools to scan every single data record.
* Performed data analysis, data validation, data cleansing, and data verification to identify data mismatch using **Relational Data modeling (3NF)** and **Dimensional Data Modeling.**
* Performed exploratory data analysis on all the features to understand feature importance and analyzed the behavior of features by using different statistical approaches.
* Studied the feature distribution with the help of Probability Density Function, Cumulative Distribution Function, Percentiles, Quantiles to draw some insights.
* Tackled highly imbalanced fraud dataset using undersampling, oversampling with **SMOTE** and cost-sensitive algorithms using **Python Sci-kit Learn**.
* Developed automated model training, testing & deployment via machine learning continuous delivery pipelines.
* Built decision tree model from the set of training data using the information entropy and the attribute with the highest normalized information gain is chosen to make the decision of credit approval.
* Used ML algorithms logistic regression, support vector machine, k nearest neighbors, Naïve Bayes, CART, bagging, boosting, ensemble learning to analyze the data based on the features selected for data-driven decisions.
* Performed text analysis on the reviews of the products using **NLP** techniques like Bag of Words, Term Frequency-Inverse Document Frequency, Word2vec, Average Word2vec with help of **NLTK** library and **Gensim** package.
* Used machine learning algorithms to forecast the company’s short-term and long-term growth in terms of revenue, number of customers, various costs, stock changes etcetera.
* Used Classified instances, **Relative Operating Characteristic curve (ROC)** and **Confusion Matrix** to find the accuracy of the models built.
* Acquired knowledge on designing, iterating and fine-tuning **neural network model’s architecture** for runtime efficiency to achieve optimal performance.
* Visualized results in python using Matplotlib, Seaborn libraries of Scikit-learn and used Tableau to create the interactive dashboards to present results for team members, management and clients.

## Binghamton University, NY Jan 2013 – Dec 2015

### Role – Junior Data Scientist

Responsibilities:

* Worked with data scientists to support model building, scoring, monitoring, and reporting.
* Co-ordinated with business users to gather business requirements and prepared the documentation for analysis.
* Identified gaps in different processes and implement process improvement initiatives across the business improvement model.
* Was a liaison between project teams, data architecture, data management, data stewardship, lines of business & the delivery/development group to align business needs with enterprise data management strategy & solutions.
* Assisted in supporting the enterprise conceptual and logical data models for analytics, operational and data mart structures using an industry standard model, where possible
* Acquired data from primary or secondary data sources and maintained databases/data systems and developed data collection system & other strategies that optimize statistical efficiency and data quality.
* Used **SQL** for creating and using Views, User Defined Functions, Triggers, Indexes and Stored procedures involving joins and sub-queries from multiple tables.
* Established relationships between the tables using primary and foreign key constraints using **SQL** triggers.
* Strong ability to Merge datasets, clean constructed datasets, produce summary statistics, conduct difference in means tests, and store all accompanying files in an organized manner.
* Load and transform large sets of structured, semi-structured and unstructured data.
* Prepared and analyzed the data includes locating, profiling, cleansing, extracting, mapping, importing, transforming, validating or modeling.
* Performed data mining, working with complex data sets, conducting multiple regression analysis and leveraging statistical tools.
* Identified patterns, data quality issues, and opportunities and leveraged insights by communicating opportunities with business partners.
* Applied linear regression to understand the relationship between different attributes of the dataset and causal relationship between them using **R**.
* Performed statistical analysis to understand the data & produced forecast trends for various categories.
* Closely monitored the operating and financial results against plans and budgets.
* Made tables, charts, and graphs to visualize analysis for reports to clients using excel and tableau.
* Involved in story-driven Agile development methodology and daily scrum meetings.

## Anthem, Inc., NY Jan 2012 – Dec 2013

### Role – Data Analyst

Responsibilities:

* Involved in creating stored procedures and **SQL** queries to import data from **SQL server** to the **Tableau**.
* Create filters, parameters and calculated sets for preparing dashboards and worksheets in **Tableau.**
* Acquired strong experience in all areas of **SQL** server development including tables, user functions, views, indexes, stored procedures, functions, joins.
* Explored dataset using various diagrams such as Histograms, boxplots, skewness in **R studio.**
* Analyzed the customer data and business rules to maintain data quality and integrity.
* Extensively created excel charts, pivot tables, functions in **Microsoft Excel** to analyze the data.
* Clean dataset by removing missing values and outliers using R studio
* Perform various mathematical functions such as max, min, log, round, sum, mean, standard deviation in R studio.
* Performed **ETL** process to Extract, Transform & Load the data from OLTP tables into staging tables & data warehouse.
* Establish credibility and strong working relationships with stakeholders and customers.
* Produce reports on ad hoc basis per requirements.
* Conduct User Acceptance Testing (UAT) for various system releases.