**MINAKSHI KESARWANI**

# PROFESSIONAL PROFILE

Passionate Data Engineer around **8 years of Data engineering/ Data science experience** in automobile, Banking, Insurance domain with excellent **team building** and **management skills**. Skilled in **Machine Learning, Natural Language Processing, Data Mining, Data Modeling, Data Visualizing, Data Management, Relational and Non-relational Database**. **Business Analytics Graduate (MS)** in data science specialization, with a demonstrated history of delivering large, complex projects. I have a passion for technology and the positive impact it creates in business transformation and process improvement.

# PROFESSIONAL SUMMARY

* **MS in Business Analytics – University of Texas at Dallas**
* Have good Data pipeline creation and Data visualization experience with **Palantir Foundry tool using apps like Slate, workshop, Contour, Taurus, Ontology Manager, Object view, Report**
* Extensively used **JavaScript** to develop dashboard and **python transform** to create Data pipeline
* Have worked on Azure platform databricks, Experience in developing Spark Application using Spark-Sql in Databricks for data extraction, transformation, Aggregation from various file format e.g txt , avro, parquet, csv for analyzing customer data and infer meaningful insight
* Utilized Spark and PySpark to streamline data processing tasks, enhance scalability, and optimize performance.
* Strong analytical skills with the ability to collect, organize, analyze, and disseminate large datasets with attention to detail and accuracy.
* Leveraged strong Python skills for data manipulation, utilizing Data Frames, APIs, and batch processing to construct robust and efficient data pipelines.
* Experienced in designing, coding, debugging, reporting, analyzing data utilizing Python worked on IDE like **PyCharm, Visual Studio Code and Jupyter Notebook**
* Expertise in **Python (2.x/3.x)** programming with multiple packages including **NumPy, Pandas, Matplotlib, SciPy, Seaborn and Scikit-learn.**
* Demonstrated skills to implement and develop machine learning or deep learning models in scikit-learn or **TensorFlow such as decision trees, linear/logistic regression, naive Bayes, SVM, PCA, Neural Network**
* Good Knowledge in **Big Data with Spark API; Spark SQL, Spark Data frame, UDF**
* Hands-on experience on various function, transformation and actions performed on SparkRDDs
* Experienced in importing data onto Hadoop Distributed File System using MapReduce and performed analysis using Hive and reported insights using Tableau 2020.1
* Worked with different File Formats like excel, **CSV, Parquet and Avro for Hive querying and processing**
* Hands on experience on visualization tools and performing Exploratory Data Analysis (EDA) **using Tableau 2020.1, matplotlib and seaborn in Python and GGPlot in R**
* Generated **Tableau visualizations** and dashboards using tableau desktop. Generated Dashboards with Quick filters, parameters and sets to handle views more efficiently
* Combined visualizations into Interactive Tableau Dashboards and published them to the web portal such as Tableau public and Embedded Tableau interactive dashboard into webpage.
* Extensively used advance chart visualizations in Tableau like Dual Axis, Box Plots, Bullet Graphs, Tree maps, Bubble Charts, Waterfall charts, funnel charts etc., to assist business users in solving complex problems
* Experience working in a Test-Driven Development environment and SDLC such as Agile (Scrum)
* Excellent communication skills, self-motivated with a high degree of attention, a good Team Player, can work independently committed to work and have sound leadership qualities along with good project managing skills

# EDUCATION

**The University of Texas at Dallas, Richardson, Texas**

*M.S., Business Analytics*  May 2020, **GPA 3.94**

**SRM University, Chennai, India**

*B.Tech. Biotechnology*  August 2011, **GPA 3.48**

# SKILL SETS:

**Programming:** SQL, Python, R, Linux, Shell Script, Javascript, VB script

**Data Visualization:** Tableau 2020.1, MatplotLib and Seaborn in Python, GGPlot2 in R, Excel , Palantir foundry

**Big Data/Cloud:** Hadoop, Hive, Spark, Pig, Sqoop, Map Reduce, Yarn, Kafka, Enterprise Data Warehousing

**Cloud:** AWS S3, EMR, EC2, Google Cloud Platform

**RDBMS:** MS SQL Server, MySQL, MongoDB

**Libraries:** NumPy, SciPy, Pandas, Scikit-Learn, TensorFlow, Keras, Pytorch, Spacy, NLTK, Vader

**Software/Tools: Palantir Foundry,** PyCharm, Jupyter Notebook, Visual Studio Code, HPQC, HP-UFT, R Studio

**Machine Learning:** SVM, Regression, Classification, KNN, PCA, Natural Language Processing

**Statistical Model**: Linear Model, ANOVA, t-test, chi-square, Hypothesis Testing

**Certifications:** Tableau Analyst Badge, Tableau Data Scientist Badge, Google Analytics IQ, GoogleAdsSearch, ISTQB Foundation level certified, Pega Certified Business Architect

# PROFESSIONAL EXPERIENCE:

**Client: Credit Suisse, Raleigh, North Carolina**  **Oct 2021 – Till present**

**Data Analyst/ Data Engineer**

**Project Description:** The project is about migrating data from legacy system to cloud and creating business analysis dashboard using visualization tool of Palantir foundry; Creation of data pipeline by writing pyspark transformation and then display crucial KPI by either using Workshop, Contour or slate visualization tool.

**Responsibilities:**

* Developed data pipeline using pyspark, Sql function to pull raw data from server and moved to pcloud after preprocessing the data; compare the resultant data with the legacy system data to identify any discrepancies or data loss.
* Implemented Linux scripts to migrate data from servers to various development environments.
* Collaborated with stakeholders to understand their requirements and Leveraged Palantir Foundry's Workshop module, Ontology, and Action feature to translate them into a comprehensive interactive analytical dashboard.
* Developed and managed a data pipeline schedule using Palantir Foundry's Monocle tool to automate the dashboard's updating process whenever a new data release occurred.
* Developed Slate dashboard and updated its card widget by using writeback dataset and designed a workshop module to enable this functionality
* Created objects using ontology manager, established relationship between those objects and used them to create object view, Slate dashboard, workshop modules. Integrated object view to display information on granular level by establishing relationship between various objects.

**Technical Environment**: Palantir Foundry (Workshop module, Ontology, Slate, Action feature) Pyspark , Python , Database, SQL , HTML, Javascript, Typescript, ALM, Unix Shell Scripting, Machine Leaning, Classification, Regression, NLP, Enterprise Data Warehousing

**Client: STELLANTIS, Aurbun Hills, Michigan**  **Oct 2020 – Sep 2021**

**Data Engineer**

**Project Description:** The project is about creating data pipeline to extract, preprocess data from source using python transform and then create interactive dashboard using JavaScript displaying important KPI which can help business partner or plant user to identify critical engine which need immediate attention. They can easily identify and trend which is present with Station and create user a communication channel to immediately notify team leader about issue.

**Responsibilities:**

* Connected with source database and created data pipeline, made node incremental or snapshot according to requirement, wrote python transform to make code reusable across all plant.
* Created Dashboard using Palantir Foundry platform, created customize JavaScript for al visualization.
* Developed interactive dashboard to get information at granular level
* Followed Agile methodology to get feedback from Business partner and enhanced dashboard functionality accordingly
* Did AWS cost optimization to bring down overall project cost
* Created and scheduled daily/weekly reports
* Developed UI for tracking issues and embedded in dashboard using JavaScript and HTML

**Technical Environment**: Palantir Foundry, Pyspark , Database, SQL , HTML, Javascript, Machine Leaning, Classification, Regression, NLP, Python, AWS, Unix Shell Scripting

**Next Capital Tech LLC, Minnesota**  **Jun 2020 – Sep 2020**

**Artificial Intelligence / Data Engineering**

**Project Description:** The project is about fetching cryptocurrency data from Binance exchange site and developing a neural Network model which can predict whether a user should sell, buy and do nothing. With this, we scraped article related to cryptocurrency from various social media platform using **Beautiful soup** Toolkit and analyzed sentiment around each cryptocurrency

**Responsibilities:**

* Extracted data from a cryptocurrency exchange site Binance, preprocessed it and design model which can predict with high accuracy whether a person should buy/sell or do nothing.
* Trained models using Google cloud platform by creating VM instance and setting up configuration such as NVIDIA, GPU.
* Performed Real time data analysis on test data set by validating it across created model which gave nearly 99 % accuracy
* Performed Web scraping using Beautiful Soup using Twitter API to extract Data related to various Cryptocurrency
* Performed sentiment Analysis on scraped data to find out Polarity, Subjectivity, and Emotion related to each cryptocurrency and overall market sentiment
* Used python Libraries such as vader, SentimentIntensityAnalyser, TextBlob, NRCLex
* Experimented with various NLP techniques (stemming, lemmatization, n-grams, TFIDF, etc.), machine learning models (Logistic Regression, SVM, Navie Bayes, etc), and deep learning models (Word2Vec, CNN, RNN); found the best model based on k-Fold cross validation scores.

**Technical Environment**: Pycharm, Beautiful Soup, API, Google cloud Platform, Google colab, NLP

**Ericsson USA, Plano Texas Jan 2020 – May 2020**

**Data Analyst**

**Project Description:** To create information pool for Ericsson employees and provide them with hands on information of various Categories, issues pertaining to reaching out to relevant teams, competitor analysis and price analysis etc.

**Responsibilities:**

* Enhanced and Designed AI assistant using NLP techniques to create smooth interactions across various stakeholders and suppliers, develop applicable python actions to return intended information
* Used Pycharm to create and execute shell script as well as to write python classes
* Created database to connect Ericsson internal site to AI Assistant
* Connected Ericsson FAQ Page with AI Assistant which will result in reduced time to gather relevant information
* Created various intents, stories and Python scripts based on the RASA Natural Language Processing t train the model so that it can generate message with approximately 80% Accuracy
* Created multiples interactive dashboards with different KPI’s to plug and play and see the data at different hierarchies
* Used all the show-me Visualizations from Tableau and created interactive worksheets, dashboards and story point to create stories

**Technical Environment:** Tableau, Windows server, SQL Server, Pycharm, Rasa NLU, Rasa Core, Shell Script

**Mindtree Ltd., Bangalore, India Aug 2011 – Feb 2016**

**Senior Software Engineer**

**Project Description:** The project assignment was about analyzing general insurance dataset and provide important insight to client which can help the business improve their marketing strategy for business growth. The assignment also included automating a banking web application which would enroll clients/customers into system and provide various banking product to them with Esign Facility.

**Responsibilities:**

* Extracted, transformed Insurance and banking data using SQL server, created complex queries using functions, view and subqueries
* Created interactive Tableau operational dashboards and stories, depicting the KPIs using calculated fields, groups, action, quick and context filter and hierarchies which helped client in improving marketing strategy-based customer segmentation resulted in 10% business growth
* Developed workbooks (with over 12 worksheets) with various functionalities and deployed them on server.
* Used functionalities like groups, sets, Actions (hyperlinks), context filters wherever required.
* Involved in Creation of tables, Manipulated the data using DDL and DML functions
* Trained and mentored 10+ team members which helped in meeting on-time project deadlines.
* Created 250+ automation test scripts and updated hybrid automation framework using HP UFT 11.5 and VB scripting, reported 50+ defects logged into HPQC, assigned to developers and conducted a walkthrough to explain the issues, retested, closed the defects and improved application functionality by 25%

**Technical Environment:** Tableau, Windows server, SQL Server, HP UTF,HP ALM, VB Scripting

# ACHIEVEMENTS

**Itelligence-Analytics-Challange-5.0: Winner Among 40 Teams –1st Position**

***(Python, Tableau)***

* Preprocessed and analyzed UNICEF data using Python, created dashboard using Tableau to visualize the insights, extracted various factor which can improve quality of kid’s life in various countries
* Identified the trend present in data and made recommendations accordingly