**Anudeep RESUME**

[**Anu.deep65@outlook.com**](mailto:Anu.deep65@outlook.com) **| 214 227 2114**

**Location- ST Louis MO**

**Data Scientist/Python | Machine Learning Engineer**

Technical professional with an extensive portfolio of projects, passionate for data with global experience around the spectrum in Data Analytics and Data Science for tackling challenging problems with specialization in Machine Learning and Predictive Analytics. A motivated problem solver with an aptitude for innovation and solutions development.

**SUMMARY**

Profound experience as a data scientist with deep understanding of technology trends with expertise in core of complex technologies.

* Experience with various ingestion techniques to bring into R, Python and Azure ML environment from different big data platforms such as a HDFS, Hadoop
* Hands on with various Data cleansing process like handling missing values by using feature selection such as a replacing by mean, forward or backward fill, removing entire rows or columns or values, removing outliers, normalizing and scaling data
* Visualized data using different visualization tools R, Azure ML and Power BI
* Created predictive model using supervised, unsupervised and ensemble machine learning algorithms
* Hands on experience in implementing classification algorithms such as KNN, Naive Bayes, Decision Trees, Clustering, Linear and Logistic Regression
* Familiar with predictive models using numeric and prediction algorithms like Support vector machines and Neural Networks, and ensemble methods like Bagging, Boosting and Random Forest to improve the efficiency of the predictive model
* Experience in extracting the data for creating value added datasets using Python, R, Azure and SQL to analyze the behavior; to target a specific set of customers and obtain hidden insights within the data to effectively implement project objectives
* Worked on KAGGLE data sets and Microsoft Azure ML predictive models as a part of Data science Boot camp
* Experienced the full software life cycle in SDLC, Agile and Scrum methodologies
* Strong conceptual, analytical, and design skills with leadership qualities. Able to work with a team or individually with excellent communication skills, and ability to meet deadlines in a fast-paced work environment

**EDUCATION**

**Master of Science** - Computer Science, 2016 | Chicago State University | 3.4 GPA

**Technical Proficiencies**

|  |  |  |  |
| --- | --- | --- | --- |
| * Python, R * PL/SQL, T-SQL | * Azure Blob, Data Lake * HDFS | * JSON, XML, XHTML * UNIX Shell Script | * Linear & Logistic Regressions, Clustering * (K-means) * Bayesian Algorithm, KNN, Random Forest |
| * C, C++, C# * Core Java | * MS SQL Server, MS Excel, Power BI * ODBC, IDE * GIT HUB | * Azure ML * KAGGLE, Tensor flow, Keras | * Neural Networks, SVM, Bagging & Boosting |

**Professional Experience**

**Questar Assessment, Inc., Apple Valley, MN Mar 2017 – FEB 2019**

**Data Scientist / Machine Learning Engineer**

**Project:** Questar Assessments Inc. is an educational assessment provider for K-12 students. There are multiple projects within the organization related to processing of student data, loading DataMart from transactional and legacy systems, creating interactive reports for districts and schools and generating data files.

* Worked with business users, business analysts, program managers, project managers, system analysts for reviewing business requirements
* Created technical specification documents based on business requirements and collaborated with BA and other software developers to discuss design and architecture
* Good experience of software development in Python and IDEs: pycharm, Jupyter Notebook.
* Developed predictive analytics models on Python for the utility of scoring essay program and slide of the reader scores of 500,000 Pre-K to twelfth-grade students of the schools in Tennessee State
* Developed tools using Python, Shell scripting, XML to automate some of the menial tasks.
* Used Python to write data into JSON files for testing Student Item level information. Created scripts for data modelling and data import and export
* Performed Data preparation on a High dimensional (Big data with large volume and variety) data sample collected from the students essay tests data using in to Python
* Data preparation Includes Data Mapping of unlined data from various formats, Identifying The missing data, Finding the correlations, scaling and removing the Junk data to further process the data for building a predictive model into Apache Spark
* Data cleaning, pre-processing, imputation, transformation, scaling, feature engineering, data aggregation, merge data frames, descriptive statistics, data visualization, score assessment mapping, reporting on Tableau dashboards
* Worked on Azure databases, the database server is hosted on Azure and use Microsoft credentials to login to the DB rather than the Windows authentication that is typically used
* Used Docker to run the local instance of the application to laptop. Run Docker app in the background to test the application and simultaneously query the local instance of the database in order to see which tables are inserted/created. Used commands such as Docker-compose up
* Built Power Shell scripts to execute Bulk copy protocol (BCP) commands to pull bulk data into database (we process around 150 Million records)
* Closely working with Document Processing and Scoring Services Expert Team to find the rule sets to build a predictive model and performed visualization for getting in depth knowledge and correlation between variables
* Processed data using Python and developed a Predictive model to predict KPI’S (Key performance indicators) such as domain level scores within ranges and Retain Ability
* Conduct exploratory data analysis using Pandas, NumPy, Matplotlib, Scikit-learn, SciPy, and NLTK in Python for developing various machine learning algorithms
* Perform data extraction and manipulation over large relational datasets using SQL, Python, and other analytical tools.
* Closely working with senior Artificial Intelligence Team to create and build a Machine learning layer in the final Product
* Trained Data with Different Classification Models such as Decision Trees, Random forest, Linear & Logistic Regression, KNN models to classify quartiles & predict scores
* Based on over all Statistics, Model Performance and Run Time decided Final Model and achieved accuracy, precision, recall in the range of 75-80 % on average for the validated models
* The program showed improvement in the essay score of 75.7% students from fall 2017 to fall 2018, the scores were improved by equivalent of 0.1 months of schooling for each book they read, 1.4 months if they participated last year as well and by 1.2 months for students with economic disadvantage
* Used Python libraries and SQL queries/sub queries to create several datasets which produced statistics, tables, figures, charts and graphs.
* Manage, develop, and design a dashboard control panel, Graphical representations, Pie Charts etc., for Program to see the Student Performance in individual reports using Django, HTML, CSS, JavaScript, and JQuery calls.
* Follow the process of updating and maintaining JIRA support ticket, Project story and its sub-tasks workflow process and communicating with ticket submitter. Maintained a track of all the loads in JIRA.
* Upload detailed documents on process flows, ETL flows, explanation of scripts used for validating data files, DataMart tables in confluence for knowledge sharing and team building.

## **Achievements**: Presented models and results to different State Clients in a conference for Business Attraction

**Technology Stack**: Python, Django, HTML, CSS, JavaScript, Shell Scripting, SQL, Visual Studios, Integration Services, Power BI, Azure ML, Tableau, Dockers

**What was the spectrum for Prediction?**

In depth analysis of data, calculation of KPIs from the performance metrics data and based on the current KPIs metrics predicting the KPI for Next Admin / Next Term

## **What was the action Taken?**

Closely worked with senior data scientists for analyzing the predictive model which was built using Machine Learning Techniques purely a Classification problem

**Health iPASS Inc., Oak Brook, IL Jan 2015 – Mar 2017**

**Data Scientist / Data Analyst**

**Project:** Load the Integrated Eligibility System (IES) data into Human Services Data Warehouse, allowing executive office of Health & Human Services to perform data analysis, generate reports, and monitor recipient eligibility.

* Participated in requirement gathering and worked closely with the architect in designing and modeling.
* Developed a fully automated continuous integration system using Git, Jenkins, MySQL and custom tools developed in Python and Bash
* Creation of Python scripts for data access and analysis to aid in process and system monitoring, and reporting
* Analyzed data and performed data preparation by applying historical model on the data set in AZURE ML
* Developed predictive models based on demographic, psychographic, econometric and statistical data that deliver insights related to member enrollment
* Performed Data cleaning process applied Backward - Forward filling methods on dataset for handling missing value
* Under supervision of Sr. Data Scientist performed Data Transformation method for Rescaling and Normalizing Variables
* Developed a predictive model and validate KNN model to predict the feature label
* Performed Boosting method on predicted model to improve the efficiency of the model
* Job scheduling, batch-job scheduling, process control, forking and cloning of jobs and checking the status of the jobs using shell scripting
* Used PyQt for the functionality filtering of columns helping customers to effectively view their transactions and statements. Implemented navigation rules for the application and page.
* Created Data tables utilizing PyQt to display customer and policy information and add, delete, update customer records.
* Worked extensively on developing a shell-script which automates the packaging of the project and deploys it from Jenkins to the production server when the project is pushed to the GitLab.
* Deployed the project into Jenkins using GIT version control system.
* Used Shell scripting for host concurrent programs and migration scripts for deployment.
* Wrote Unix Shell Scripts, undertook Code Optimization and Performance tuning of the application.
* Presented Dashboards to Higher Management for more Insights using Microsoft Power BI

**Technology Stack**: Python, Python libraries such as NumPy, Pandas, SQL, Alchemy, Shell Scripting, JavaScript, MySQL, GIT, PyQt, Microsoft Power BI, Azure ML

## **What was the prediction?**

To investigate, rationalize, and recommend optimal approaches to positively impact engagement and health outcome key performance indicators, to create a predictive model to support consumer engagement within our different product offerings based of model and past data

## **What was the Action Taken when a system was built?**

A system was successfully created on the past data, medical history, health activity of customers and runs model against the historical data and get predicted label if customers are eligible for product offer or not, on the base of that send them offers

## **What was the action Taken?**

Customers should get actual offer they like and accept it which increases the numbers of consumer engagement that leads to profit the company

(Note: gap due to obtaining Master’s degree)

**Cloudizen India Pvt Ltd., Hyderabad, India Sep 2013 – Nov 2014**

**Database Developer**

**Project:** The project was about the health care financial steam line management system. Main object of this project is to identify the recipients eligible for insurance claims to submit its Quarterly data and calculate quarter’s financial ratios.

* Collected business requirements from users and translated them as technical specifications and design documents for development.
* Used Entity Relationship Diagrams and Data mapping for better understanding of dataset, Data Modeling
* Responsible for creating ETL packages using SQL server integration services.
* Developed logging for ETL load at package level and task level to log number of records processed by each package and each task in a package using SSIS.
* Responsible for Deploying, Scheduling Jobs, Alerting and Maintaining SSIS packages.
* Designed SSIS packages using Lookup Transformations, Merge Joins, Fuzzy Look Ups, and Derived Columns with Multiple Data Flow Tasks.
* Created complex Stored Procedures, Triggers, Functions, Indexes, Tables, Views and other SQL joins for applications following SQL coding standards.
* Performed Statistical Analysis and Hypothesis Testing in Excel by using Data Analysis Tool
* Supported and maintained the existing SSRS reports and responsible for source code fixes
* Experience in all phases of Software Development Life Cycle (SDLC) such as Requirement gathering, Analysis, Development, Testing, Implementation and Support.
* Developed internal dashboards for the team using Power BI tools for tracking daily tasks.
* Created the reports and data extracts from SQL server using Excel pivot as per user requirements

**Technology Stack**: T-SQL, MS SQL Server 2010/2012, SQL Server Integration Services, SQL Server Reporting Server, Power BI, Power Pivot, MS Excel, Regression

**KAGGLE**

**State farm Distracted Driver Competition**: Given dataset of a 100,000 driver images, each taken in a car with a driver doing something in the car (texting, eating etc.) with the goal of predicting the likelihood of what the driver is doing. After first utilizing Tensor flow, Keras, and AWS p2.x16large instance to pre-process images, a pre-trained VGG-16 model, where the last convolutional output of VGG-16 is fed back into the model, was utilized. A global average pooling layer & a fully connected layer were added and uses the activation, softmax. An SGD optimizer was incorporated with a slower learning rate and momentum of 0.9. Model Version 3 resulted in the best score, 1.26, and equated to a rank of 615/1440 (a top 40% finish) on the Kaggle Public leaderboard.