Ashwith Gundu

(660) 528-0658

gunduash@gmail.com

DATA SCIENTIST

|  |  |  |
| --- | --- | --- |
|  | Summary |  |

* 7+ years experience in Machine Learning, Data mining, Data Architecture, Data Modeling, Data Analysis, NLP with large data sets of Structured and Unstructured data, Data Validation, Predictive modeling, Data Visualization, Text mining to transposing words and phrases in unstructured data into numerical values.
* Worked with complex applications such as Python libraries to develop neural network, cluster analysis.
* Expertise in all aspects of Software Development Lifecycle (SDLC) from requirement analysis, Design, Development Coding, Testing, Implementation, and Maintenance, followed Agile methodologies
* Experience in designing stunning visualizations using Tableau software and publishing and presenting dashboards, Storyline on web and desktop platforms.
* Hands on experience in implementing Naive Bayes and skilled in Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, neural networks, Principle Component Analysis and knowledge on Recommender Systems.
* Experienced with machine learning algorithm such as logistic regression, random forest, XGboost, KNN, SVM, neural network, linear regression, lasso regression and k-means.
* Developing Logical Data Architecture with adherence to Enterprise Architecture.
* Adept in statistical programming language like Python including Big Data technologies like Hadoop 2, HIVE, HDFS, MapReduce, and Spark.
* Experienced in Spark 21, Spark SQL and PySpark.
* Done Clustering, regression and Classification using Machine learning library MLlib(PySpark).
* Skilled in using Numpy, NLTK and Pandas in python for performing Exploratory data analysis.
* Very good experience and knowledge in provisioning virtual clusters under AWS cloud which includes services like EC2, S3, SageMaker and EMR.
* Good understanding of Teradata SQL Assistant, Teradata Administrator and data load/ export utilities like Fast Load, Multi Load.
* 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.

|  |  |  |
| --- | --- | --- |
|  | Technical Skills |  |

|  |  |
| --- | --- |
| Programming | Python (NumPy, Pandas, Scikit-Learn, Matplotlib, Seaborn), SQL, PySpark, Scala, C++, Java, Javascript, HTML, CSS |
| Analytics and Visualization Tools | Tableau, MS Excel |
| Statistical methods | ANOVA, ARIMA, Regression Analysis, Hypothesis Testing, Time Series, Regression Models, Splines, Confidence Intervals, Principal Component Analysis and Dimensionality Reduction |
| Amazon Web Services | S3, EC2, EMR, Cloudformation, SageMaker and Rekognizer |
| Machine Learning Algorithms | Logistic Regression, Linear Regression, Decision Tree, Random Forest, Gradient Boosting, Nearest Neighbor Classifier, Weight of Evidence & Information Value (WOE & IV), K-means clustering, Affinity Propagation, Principal Component Analysis, Support Vector Machines, Naive Bayes, Auto Regression, Lasso Regression & Moving Averages. |
| Deep Learning | TensorFlow, Keras |
| Big Data Tools and Technologies | HDFS, PIG, MapReduce, HIVE, SQOOP, FLUME, HBase, Storm, Kafka, Elastic Search, Flume, Storm, Kafka, Elastic Search, Redis, Flume, Scoop. |
| Other Tools | Jupyter Notebook, Git Version Control, IPython Notebook, Unix, Visual Studio Code, Net beans, Visual Studio code |

|  |  |  |
| --- | --- | --- |
|  | Professional Experience |  |

|  |  |  |
| --- | --- | --- |
| Anthem, Inc., Norfolk, VA | *February 2017 - Present* | |
| Role: Data Scientist | |  |

Description: The goal of the company aims at developing the technology, solutions and programs that give consumers greater access to care, working with providers to ensure a quality health care experience for consumers and easing cost challenges by advancing affordability in the health care industry. This job involved creating statistical machine learning models for generating report on individual customer health care system followed by future insurance claims based on current customer information with multiple analytics.

⇒ 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 16.78%.
* Used Pandas, NumPy, Scikit-learn in Python for developing various machine learning models such Random forest and step-wise regression.
* 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.
* Worked on NLTK library in python for doing sentiment analysis on customer product reviews
* Worked on migrating MapReduce programs into Spark transformations using Spark and Scala, initially done using python (PySpark).
* Developed various Spark applications using Scala to perform various enrichment of these click stream data merged with user profile data.
* Developed highly optimized Spark applications to perform data cleansing, validation, transformation and summarization activities
* Data pipeline consists Spark, Hive and Sqoop and custom build Input Adapters to ingest, transform and analyze operational data.
* Created and maintained reports to display the status and performance of deployed model and algorithm with Tableau.
* 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.

|  |  |  |
| --- | --- | --- |
| Carnival Corp, Miami, FL | *September 2014 - February 2017* | |
| Role: Data Scientist | |  |

Description: It is the world’s largest leisure travel company, provides travelers around the globe with extraordinary vacations at an exceptional value. In this project we built various machine learning model to help the organization in determining the good customer base, future travel suggestion based on current customer information and other analytics.

⇒ Responsibilities

* Developed applications of Machine Learning, Statistical Analysis and Data Visualizations with challenging data Processing problems in sustainability and biomedical domain.
* Compiled data from various sources public and private databases to perform complex analysis and data manipulation for actionable results.
* Designed and developed Natural Language Processing models for sentiment analysis.
* Worked on Natural Language Processing with NLTK module of python for application development for automated customer response.
* Used predictive modeling with tools in Python.
* Worked with the Spark for improving performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Spark MLlib, Data Frame, Pair RDD's, Spark YARN.
* Applied concepts of probability, distribution and statistical inference on given dataset to unearth interesting findings through use of comparison, T-test, F-test, R-squared, P-value etc.
* Applied linear regression, multiple regression, ordinary least square method, mean-variance, theory of large numbers, logistic regression, dummy variable, residuals, Poisson distribution, Bayes, Naive Bayes, fitting function etc to data with help of Scikit, Scipy, Numpy and Pandas module of Python.
* Applied clustering algorithms i.e. Hierarchical, K-means with help of Scikit and Scipy.
* Worked on Clustering and classification of data using machine learning algorithms.
* Used Tensor Flow machine learning to create sentimental and time series analysis.
* Developed visualizations and dashboards using ggplot, Tableau
* Built and analyzed datasets using Python, Seaborn, and MatLab
* Applied linear regression in Python and SAS to understand the relationship between different attributes of dataset and causal relationship between them
* Designed and implemented a probabilistic churn prediction model with 100k 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.
* Pipelined (ingest/munge/clean/transform) data for feature extraction toward downstream classification.
* Expertise in Business Intelligence and data visualization using Tableau.

|  |  |  |
| --- | --- | --- |
| Pfizer Inc. Collegeville, PA | *July 2013 - July 2014* | |
| Role: Data Analyst | |  |

Description: Pfizer is one of the world's premier innovative biopharmaceutical companies, collaborate with health care providers, governments, and local communities to support and expand access to reliable, affordable health care around the world. The goal was creating customer profiling models and customer value analysis to improve health and well-being at every stage of life. 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.
* Worked on preprocessing of data which involves collecting, formatting, cleaning, aggregation, segregation of large volume of data and finally sampling data from it for performing statistical evaluations further inferred valuable conclusion from data.
* Developed Natural Language Processing to automate the classification of customer incident queries into levels of classes to improve the customer services
* Implemented number of customer clustering models and these clusters are plotted visually using Tableau legends for the higher management.
* 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 models (CountVectorizer, IDF) to prepare the data.
* 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.
* 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 parameter tuning and model evaluation techniques Confusion Matrix, Cross validation. Customer Profiling models using K-means and K-means++ clustering algorithms to enable targeted marketing.
* Implemented dimensionality reduction using Principal Component Analysis and k-fold cross validation as part of Model Improvement.
* Worked with data visualization tools in python like matplotlib.

|  |  |  |
| --- | --- | --- |
| Pfizer Inc. Collegeville, PA | *May 2011 - July 2013* | |
| Role: UI/UX Developer | |  |

Description: Initially collaborated on developing a web application where my major contribution to the project was to work on User Experience Design part.

⇒ Responsibilities

* Provided design expertise to the organization and work directly with web development and production teams
* Worked closely with the Product Manager and team leads to ensure we were developing world-class applications with UX/UI design expertise
* Collaborated with other designers, user researchers, game designers, engineering teams, and business/marketing stakeholders to prioritize UX activities throughout the game/application development life-cycle and deliver high quality experiences on schedule
* Worked in graphic design area, an excellent eye for typography, clean layout, purposeful color, and attention to detail. Developed deep appreciation for simple, fun, intuitive and usable interfaces
* Worked on in-house UI tools and scripting language where problems were solved using a combination of JavaScript, JSON, and JQUERY
* Having knowledge in HTML, CSS, browser compatibility and web standards for interactive prototypes, plus Adobe Creative Suite (primarily Photoshop) or similar tool for wireframe and static visual designs.