

Avinash Baggam

Data Scientist

Address

Andhra Pradesh,
Vizianagaram, 535004

Phone

+91 8978245068

E-mail

avinash0326@gmail.com

LinkedIn

<https://www.linkedin.com/in/avinash-baggam-516670a3/>

Python

Pandas

Natural language processing
(NLP)

Flask

Data analytics

DataBricks

Recommender System

TensorFlow 2.0

Keras

MQTT

Results-driven computer science professional with good technical skills, a firm grasp of business needs and understanding of user requirements gathering. Able to assess risks, troubleshoot problems and conduct tests. Excellent communication and planning abilities

SUMMARY

- Having 7.2 years of total experience in IT field and 6 years in applying Deep Learning algorithms to Computer Vision, NLP techniques, Python Automation Tasks
- Hands-on experience with NLP, mining of structured, semi-structured and unstructured data using pyspark
- Python Automation : Had deployed few Scripts on Prod Server and tried to reduce the resource Manual work. For this Used Pyspark with Pandas
- Extract insights and actionable recommendations from large volumes of data
- Adapt and adhere to industry standards while working with multicultural teams
- Excellent communicator with strong leadership skills and ability to work independently
- Experience working with Agile development methodologies like Jira
- Achieved 82% accuracy by predicting Car Exterior Damage with trained on 1000 annotated images using Mask R-CNN architecture to speed up insurance claims
- Experience with tools like Control-M
- Developed and implemented algorithm to Parse resume using Tesseract and EaseText

Work History

Data Scientist

Thakral one Solutions India PVT.LTD

2022-01

Present

- Experience working with ChatGPT . Developed tool for the customer Language Modeling and BERT to implement algorithms on the same
- Worked on LLM Models for the customer text representations and which inputs the text to an embeddings model and extracted the numerical representation of this text.
- Pulling Data Incidents from CGR Database
- Connect to datalake and fetch the columns. Used multithreading to pull the data from datalake as incidents data were too large to handle
- Used Service Now tool API tool for Tasks Assignment Automation to the respective Resources Teams Based on Roaster Shifts timings (Start_time to End_Time)
- Auto Categorization of Snow Incidents Queries by considering GOS Connectivity with Kerberos Authentication
- Extensively worked on Control- M Job and also Manged to Monitor Python code Automations whatever the code has been deployed to Prod Parallel Servers
- Extensively worked on HTML Tabular Mail Format For this I have converted the pdf document to the image since tesseract is an OCR engine that works fine on image data.

Snow incident Auto Closure/Auto Transfer :

Auto Transfer : Developed tool to pull incident L queue raw data from CGR DB, also to filter only Incident data and map it to corresponding SNOW queues and also to transfer the incidents from one queue to another queue

Auto Closure : Developed tool which enables auto resolving of snow incidents. Automatic Resolution ensures timely and correct closure of Incidents

- Based on Shift Window Timings generated Handover email by connecting to Oracle DB Queries and Schedule Job on daily basis in Control-M
- Extensive experience in dealing with OracleDB including Joins, UnionAll, Union , indexes, data import/export

Senior Python Developer

GalaxE.Solutions India PVT.LTD

- Familiarity with creating and training Machine Learning algorithms (e.g. Recommendation engines, forecasting, clustering, classifiers, etc) Using Python Language
- Maintain weekly/monthly/quarterly metrics for projects and monitor products market share using pie chart
- Extensive experience in dealing with Relational Database Management Systems, including normalization, stored procedures, constraints, querying, joins, keys, indexes, data import/export, triggers and

2020-08

2022-01

cursors

- Published dashboards on Tableau public
- Interact with Business users and understanding their requirements
- Ability to use custom SQL for complex data pulls
Extracted data from the database using SQL procedures and create data sets for statistical analysis, validation and documentation.
- Worked on Web Scraping to access and extract large amounts of text data
- Recommend ways in which the business can apply data to business problems
- Experience using Docker Container
- Hands on Experience using Big Data Hadoop, Apache Hive
- Designing, creating and publishing customized interactive reports, dashboards, and scheduling using Tableau Server
- Created Tableau scorecards, dashboards using bar graphs, scattered plots, stack bars geographical maps, Gantt charts using show me functionality

2019-09
2020-07

Python Developer

Indium Software India PVT.LTD

- Familiarity with Creating and training Machine Learning algorithms(e.g., Recommendation engines, forecasting, clustering,classifiers,etc)
- Created complex stored procedures, triggers, cursors, tables and views, and other SQL joins and statements for applications
- Extensive experience in dealing with RDMS, Tensorflow, R-CNN, Mask RCNN Image Classification scikit-learn, openCV MQTT Pandas Classification and Regression Techniques

2016-02
2019-08

Python Developer

Miracle Software Systems India PVT.LTD

Experienced in designing and development of Tableau visualizations and create business requirement documents and plans for creating dashboards.

- Worked with different set of algorithms like RandomForest , XGBoost , Logistic Regression ML algorithms
- Created action filters, sets, parameters and calculated sets for preparing worksheets and dashboards in PowerBI

Certifications

IBM SPSS Modeler Professional &
Kore.Ai (Beginner and Advanced)

Education

2011-06 -
2015-04

B.Tech: Computer Science And Engineering

Miracle Engineering College (Miracle Educational Society Group of Institutions)

Platforms

Microsoft Azure

- Azure Cloud Functions, Azure SQL Server, Azure Databricks

Amazon Web Services (AWS):

- AWS Lambda, Amazon DynamoDB, Amazon S3,

Amazon EC2

BI Tool:

- Tableau

Projects

PROJECT	Healthcare Project - Databricks
TECHNOLOGIES	Microsoft Azure Databricks
ROLE	Senior Developer - Databricks
LANGUAGE/ ENVIRONMENT	Azure SQLDB, Azure ML, PySpark

Description:

The objective of the project is to identify the Formulary Solution based on Drug Data and also provide Feasible Solution for Drug Tier automation using Azure Databricks Clusters without any human intervention. Need to enhance data by simply turning on Trigger Based Jobs in Azure Databricks, continuously Stream the data based on Kafka topics for the Transformation Logic notebooks and Kafka Streaming tables as well where we have formulary historical and incremental load data (Attunity). Getting the data from OracleDB source table and use Databricks middleware(use Azure ML) and load the data into AzureSQL canonical DB form. Developed python Random Forest model in Azure ML to predict risk analysis of Patient.

ROLES & RESPONSIBILITIES:

- Controlling Project Management and Collaboratively worked together within the team

- Have Taken Oracle DB Tables as a Source for validating Data which require access to Formulary Data. Need to validate libraries for setup and Transformation and schema handling through Azure Databricks
- Have Developed Python Code based on Business logic in Transformation Notebook and streaming notebook, stored the raw data from kafka topics in Delta Lake Storage. Used tools for connecting azure servers like Nginx, Gunicorn
- Analysis on Canonical Azure DB for the different tables and logic to be implemented to push the data to Azure Delta Lake (Canonical) which is said to be Formulary Canonical Data Model (Delta Lake)
- Created the Stored Procedure to the Respective Tables in AzureDB SQL Server for validating the incremental data in Databricks Notebook
- Based on cluster it will trigger the notebook job where the data automatically pushed to Final Formulary Canonical Data ML Model (Azure SQL)
- Reduced risks processing time by 40% and claim bills costs up to 25%

PROJECT	Real Products Review and Campaign Analysis
TOOLS	NLTK
ROLE	Data Scientist
LANGUAGE/ ENVIRONMENT	Pyspark, Django, AWS, MySQL, Tableau

Description:

The objective of the project is to identify the spam reviews among the user reviews, predict the spam probability for every review and determine the attributes influencing as Spam using machine learning classification algorithms

ROLES & RESPONSIBILITIES:

- Data injected from MySQL Server into csv files and then processed in Python
- Pre-processed the data using python, performed Text Normalization and created the tf-idf matrix.
- Built the model using SVM and Random forest on the training data For feeding test data, web portal is created using flask application and deployed on AWS EC2
- Analysed the data to enable consumers make data-driven decisions about products
- Product Genius - implements its own rating system, using Bayesian averaging, to create a robust metric to the consumer instead of product's simple average rating
- Uses ML to extract the most relevant terms from positive, negative reviews of a product
- Build Real Time Tableau Dashboard and analyse the trends of the products by creating graphs

PROJECT	ML building tool (Customer Churn Prediction)
TOOLS	Python Tabula, Camelot
ROLE	Data Scientist
LANGUAGE/ ENVIRONMENT	Django, MySQL, Tableau, Web3JS

Description :

The objective of the project is to scrap the data from various sources like from website, PDF to save the data in specific format

ROLES & RESPONSIBILITIES:

- Feeding front end application with web links as input
- Pre-processes respective methods based on the type of data contained in the link using Hadoop
- Raw data in text files is tokenized, transformed, cleaned and ingested into structured text files using Pyspark. Used with tools like Nginx, Gunicorn for connecting AWS cloud
- Performing necessary data cleansing such as junk characters removal, replacements
- Analyze the data by Creating detailed level summary report using Trend lines, Statistics, log axes, groups, and hierarchies

PROJECT	Coke Image Recognition Data Analysis
TECHNOLOGIES	Tensorflow, python-OpenCV, Tableau
ROLE	Data Scientist
LANGUAGE/ENVIRONMENT	Flask, AWS, MySQL, Docker Container

Description:

Created Basic UI using Flask and for added enhancement doing analysis keeping the data as a source

ROLES & RESPONSIBILITIES:

- Automated the scraping tasks on AWS Ubuntu instance using Crontab.
- Implemented Data Validation techniques to verify the authenticity of the data.
- Storing and organizing data in AWS MySQL database
- Performed data wrangling, data cleaning and exploratory data analysis on the collected data using Python.

- Designed the model to identify whether the given image is bottle or tin by training the model with different bottle images
- Find the number of bottles/tins of each beverage product in the rack leaving void spaces. Developed modules to transform unstructured information in the form of tables
- Generated report file to give the row-wise counts of bottles and tins
- Customer can view the data on Tableau for increasing sales promotion
- Conceptualized the real-time shopper coke merchandise in retail stores by analyzing demand

PROJECT	Handwritten data detection from Images
TECHNOLOGIES	Python, Tableau, Flask, Web3Js
ROLE	Data Scientist

Description:

Retraining handwriting recognition model by feeding more images. Used CNN model to recognize the table and chunks of the tables.

ROLES & RESPONSIBILITIES:

- Aggregation of empty labels and filled columns using subsetting
- Classifying numbers using Handwritten number recognition
- Created the Django Application UI for to upload images as model has to identify the data inside each table and put in json and csv format
- Keeping the data secure and analyzed the reports on Tableau Dashboard

