Shiva Santosh

Education:

- M.C.A Sastra University 77%
- B.Sc Electronics & Tech. Loyola Academy 79%
- Maths, physics, chemistry junior college 77%

Area of Expertise:

- Data Science
- Machine Learning
- Deep Learning
- Data Visualization
- Computer vision
- SQL
- Pyspark

Certification:

- (AI-900) Microsoft Azure AI Fundamentals
- Coursera Machine Learning Stanford University
- Master Course in Tableau 10 & 2019 for Business Intelligence
- Machine Learning, Data Science and Deep Learning with Python

Intro: A Data Scientist with 7+ years of experience, passionate about data driven strategies, impactful experiments with organized technological stack.

WORK EXPERIENCE

2021 - PRESENT

• 2021 - Present DATA SCIENTIST

Us Tax Analytics ML, Shell Oil & Gas Bangalore, India Working for Shell Oil and Gas client under data science CC as a Data scientist to deliver high value Cognitive solutions. Incorporating state of the art cloud tech in data science like Azure data factory, Azure Data bricks, ML-flow, DevOps, Model Productionisation.

2015 - 2021

TATA CONSULTANCY SERVICES

• 2019 - 2021 DATA SCIENTIST

AGILE COE GRP, TCS Hyderabad, India working as a full time Data scientist for CoE group with divergent clients and problem statements

 2018 - 2019 JR.DATA SCIENTIST & DATA ANALYST IOT IOU, TCS Chennai, India
 Worked with TRDDC as a Jr Data Scientist to improve over all Efficiency of a GTCC combine cycle power plant possessed by a Japanese client.

- 2017 2018 AI & ML Developer / EXPERT IGNITE LABS, TCS Chennai, India
- 2016 2017 *IOT DEVELOPER*

IGNITE LABS, TCS Chennai, India • worked as full time IoT developer, Developed few • projects based on Robotics and Embedded Technology. • And also worked as a trainer of Python, Front end • Technologies for 2 batches of 24 each.



Secondary



Achievements and Awards:

- VPCIO Awards Us Tax Analytics ML
- Raising Star Shell.AI Awards
- Innovation Pride Award CBO Technology and Innovations
 CoE
- Ignite AI Hackathon 2.0 Winner
- On the Spot Award Ignite

SOME PROJECTS

I HAVE WORKED ON

Project Description	Technology Stack
<u>US Tax Analytics - ML Anomaly Prediction:</u> To identify Tax Errors (i.e. Anomalies) by leveraging Machine Learning Classification Model. This will help detect additional Tax Anomalies/Errors. Direct bottom line benefits by initiating faster refunds / payments, The digitalization helps overall efficiency of several days of manual efforts needing human judgement/cognition which would be possible in few seconds	Pyspark, Azure Adf, Machine Learning, Databricks, Classification Algo, ML-flow, Python, SQL, DevOps, PowerBi
US Tax Analytics - ML Misbehaving Dimensions: To analyze and predict which dimensions within the data set contribute to maximum number and value of Anomalies (Tax errors) based on Frequency, Total Amount, Total Tax Amount and Total Gross Amount. This functionality will give insights around combination of which fields (data points – ex Certain vendor material jurisdiction combinations etc.) have contributed to maximum number of Tax errors and maximum value of tax errors.	Pyspark, Azure Adf, Statistical Analysis, Databricks, DBFS, Python, SQL, DevOps, PowerBi
US Tax Analytics - ML Top Level Groups: To analyze the data and identify similarities / affinity between the data records (transactions) from Anomalous dataset and show the relevant clusters (groups) so that entire data can be categorized against the groups and relevant decisions / insights can be drawn. This functionality will aid the current validation rules in HANA (which has a preset list of clusters used for anomaly identification) by prescribing new clusters (cluster = group of data records which had data fields that have contributed to maximum anomalies).	Pyspark, Azure Adf, Machine Learning, Databricks, Clustering Algo, Pyspark K means Clustring, ML-flow, Python, SQL, DevOps, PowerBi
US Tax Analytics - ML KeyWord Extraction: Unsupervised automatic keyword extraction method that identifies the most relevant keywords in a text by using text statistical data from single texts. Extracted keywords against clusters are considered to enhance or create new rules for rule based engine. This technique does not rely on dictionaries, external corpora, text size, language, or domain, and it does not require training on a specific set of documents.	Pyspark, Azure Adf, Machine Learning, Databricks, NLP Algo, Yake Model, ML- flow, Python, SQL, DevOps, PowerBi
Multimetric correlation of infrastructure Metrics: Correlating different Infrastructure Metrics to pre-empt Infrastructure Incidents. Various Infrastructure Metrics like CPU, Memory, Disc etc are recorded using Prometheus. Future Metrics are predicted using XGBOOST Algorithm. Pearson Correlation Co-efficient is used to find correlation among various metrics before predicting. Also implemented single input multiple output technique, Predict infrastructure behavior when workload increased, and provide recommendation through what-if analysis.	Machine Learning (ML), Predictive Analytics, Regression Algorithms, Python, XGBOOST, Jupyter Notebook, Jmeter
GTCC power plant Model: Worked with TRDDC as a Jr Data Scientist & Data analyst to improve over all Efficiency and model a GTCC combine cycle power plant, used to build a digital twin. Random Forest ML model is build and delivered to mimic power plant to test and improve over all Efficiency of real system.	Python, Jupyter Notebook, predictive analytics, EDA, Sklearn, Machine Learning Algorithms
Medical Image Classification (Pneumonia Detection): For detecting the cases of pneumonia from the X-ray images as early as possible with higher accuracy. Deep learning based solution takes medical images as input, process and generate results with over 98% accuracy at the edge. The solution can work without internet connectivity and generate results at lower latency	Python, Flask, Keras, Tensorflow, Deep learning model - VGG16, GCP Cloud ML service, Intel NUC
Visual inspection and Defect detection: Unet architecture based image segmentation model to detect defects and damages from the given image and live cam feed	Python, Keras, Flask, Tensorflow, Deep learning model Unet achitecture, Deep Learning (DL), Image Analysis