Nitesh Kumar Gupta *Data Scientist*

✉ guptanitesh2711@gmail.com 📞 9121499384

📍 Hyderabad,500037 in https://www.linkedin.com/in/nitesh-kumar-gupta-6847b9218/

○ https://github.com/niteshgupta2711

Data Scientist with 1 year of internship experience in Data analytics, deep learning, big data, and machine learning. I'm looking for a challenging position in a reputable company so that I can use my skills and knowledge, further my education. Good at navigating challenging projects and expressing requirements to other team members. creative and analytical problem solver. a capable team member.

💼 Professional Experience

Data Science Intern, *Ineuron* ⧉   11/2021 – Present
Bengaluru
- Developed Advanced Image Extractor jobs with Celery Worker, Celery Beat for job scheduling, and Django as a web framework. AWS S3 for storing large amounts of images and generating zip file links.
- On a tenant-based Hadoop Ecosystem and Databricks, developed a Py-spark application (Movie Analytics) (RDD, Spark SQL, Spark data frame)
- Used MLflow to create an end-to-end Chest X-ray classification system for tracking experiments and serving models.
- Developed a proof of concept for a hyper personalised reinforced-based recommendation system was created. Curently working on a research paper about it.

Data Analyst Intern, *Infinity Learn* ⧉   12/2021 – 02/2022
Madhapur,Hyderabad
- Built Power BI dashboards for product managers for constructing strategies.
- Built Real time Dashboards for diverse groups within side the company.Built KPIs for a variety of company products.
- Part of 3-4 member Data Science Team.
- Part of the beginning segment of the Data Science Team.

🎓 Education

BTech, *Institute of Aeronautical Engineering* ⧉   07/2017 – 08/2021
8.32 CGPA | Mechanical Engineering   Hyderabad

📁 Projects

Hyper Personized Reinforced based recommendation System, *Prototype completed* ⧉   Present
Writing Research Paper

Image Classification (Chest XRay), *End-to-End Project using MLflow* ⧉
- **Problem Statement** : To create an entirely redundant system In order to classify medical X-ray images to categorise into 5 classes.Aids in the diagnosis of chest diseases in patients and determines whether a disease is relevant to a patient or not.
- **Algorithm** : CNN (Efficient Net Architecture)
- **Technology** : Tensorflow, MLflow, tensorflow input data pipelines
- **Roles and Responsibilities** : Business Understanding, Strategy and Design, Writing Input Pipeline , Data Collection, Model Building , Experimentation .

CI/CD/CT Kubeflow (TFX), *Cover Type Classification problem* ⬀

- **Problem Statement** : CI/CD/CT achieved using Kubeflow in GCP. High end compute and storage capabilities to enable tensorflow keras tuner jobs.Model monitoring with monitoring service in GCP, Experiment tracking, and Model serving end point with GCP.
- **Algorithm** : Deep wide ANN(classification)
- **Technologies** : GCP, Gitops, Kubeflow, tensorflow, Keras tuner
- **Roles and Responsibilities** : End to End to Automation for Experimentation and Model Analysis

Movie Analytics, *Analytical Load with Pyspark* ⬀

- **Problem Statement** : Using Pyspark to analyse data at scale (RDD, SparkSQL, Spark DataFrames).Taking advantage of the ease with which distributed in memory processing capabilities can be used for analytical processing
- **Algorithm** : Spark Distributed Processing
- **Technologies** : Pyspark, Databricks, hadoop
- **Roles and Responsibilities :** Data Ingestion, Strategic Analytical queries using RDD, SQL, DataFrames
- **Detailed Project Report** ⬀

Advanced Image Extractor Job, *Celery job Scheduling* ⬀

- **Problem Statement** : To download images in bulk and schedule this job for anyone. One can download any number of images of any required size for experimentation.
- **Algorithm** : selenium web Automation in python
- **Technologies** : python, selenium, celery, celery-beat, Django, Redis
- **Roles and Responsibilities** : Job Flow Design, Components selection for job Scheduling
- **Detailed Project Report** ⬀

⊚ Skills

Data Science
NLP | Computer Vision | Problem Solving | Data Analysis & Visualization | Web Scraping | Big Data | DL | ML | RL

Deep Learning
ANN | CNN | RNN | Supervised Learning | Unsupervised Learning | Reinforcement Learning

Mathematics
Inferential Statistics | Descriptive Statistics | Linear Algebra | Calculus

Big Data
Hadoop | Hive | spark | Big Query/ML

Programming Languages
python | scala

Cloud / Devops
Linux | GCP | MLops | Kubernetes | Docker | git

Machine Learning
Sklearn | Matplotlib | Data manipulation | Numpy | Pandas

Analytics
SQL | NoSQL | BI tools | statistical modelling | Excel

⊛ Tools and Frameworks

Tensorflow
ANN | CNN | RNN | Transfer Learning | Reinforcement Learning | Input Data Pipelines

MLops
DVC | MLflow | TFX

Web Framework
Flask | Django

BI Tools
Tableau | Power BI | Data Studio

Big Data
HDFS | Hive | spark |