Harshita Agarwal

harshita.ag897@gmail.com | +91 9870231462 | Ghaziabad, India
www.linkedin.com/in/harshita-agrawal-05402a129 |https://github.com/harshita897

Summary:

Data Scientist with 1.7 years of broad-based experience in building data extensive end-to-end solutions and Master's Degree (M.Tech.) in Data Analytics. Proficient in predictive modeling, data processing, Machine Learning algorithms, Image processing including Python language. Capable of developing highly adaptive & diverse AI/ML solutions to translate business and functional qualifications into substantial deliverables.

Technical Skills:

- **Programming:** Python, SQL, HTML
- **Web Scraping Tools:** Selenium, Playwright, brave
- **ML Libraries:** Scikit-learn, Pandas, Numpy, OpenCV, Pillow, Matplotlib
- **Deep Learning:** Neural Networks(RCNN, CNN), TensorFlow, Keras
- **NLP Libraries:** NLTK
- **IDE/Environments:** Visual Studio Code, Anaconda, Jupyter Notebook, Spyder, Filezilla
- **Reporting Tools:** MS Office (Word/Excel/PowerPoint)
- **Cloud Platforms:** GCP(Basics)
- **Operation Systems:** Windows, Linux

Work Experience:

- **Skribe, Delhi, India** ***(May 2022 – Present)***

  Senior Data Analyst

  Responsible for extracting many kinds of data from websites and making them automated to get the refreshed data with **BeautifulSoup** and **Selenium**.

  - Extract data from Government websites:
    - By using **Selenium** and proxies, we extracted all the data to make a product for our organization which is going to be live soon and the scripts are automated so that whenever URLs will be hit, data will be refreshed.
  - Instagram Web Pages:
    - By using **Instaloader** , **Instagrapi** extracted data and except these two, extracted data without using access tokens of any handle, got a chance to extract the data of influencers (data points as in charts with canvas , their insights, their general information, biography).
    - By Using **Instagram Official Web** , Extracted data of influencers, which was an achievement for me and i designed a bot for that to build a product for our company in which each and every detail of user will be there such as : **title, Bio, followers, followings, charts** etc.
  - Automated Crawlers:
    - By using different methods (**Selenium , BeautifulSoup, Proxy and many others**) build crawlers means an automated way to extract the data from different websites.
    - Worked with CNBC, News18 etc. websites to extract their articles with each and every data point.
    - Extracted Journalist Name and their available information as their social media URL or if any present other URLs, their latest article or when they wrote last time etc.

- **Ripik, Noida, India** ***(Mar 2022 – May 2022)***

  Analyst - Data Scientist (Part Time)

  Accountable for **developing** use cases using **Data Science** and **Machine Learning** for the offering in terms of **automation** and delivering the **end-to-end** solution by having **regular sessions** with **clients/stakeholders** to understand **business challenges**.

  - Predictive Analytics/Machine Learning:
    - **Collected data** from client's shared drive folder and **analyzed** the data to get actionable insights & make **informed decisions** about furnace energy.

- **Tooli Quality Analyst, Gurugram, India** ***(Dec 2021 – Feb 2022)***

  Computer Vision Engineer

  - Find the direction of camera:
    - Worked on **Real Time Data**(Scanned room).
    - Implemented a **logic** by which the system was able to find the direction of camera using given random values.
    - **Optimized** Models to achieve accuracy of 99.4% and up to 10MSec time Consumption.
  - Find shapes from a point cloud:
    - Worked on **Real Time Data**(Scanned room).
    - Identify shapes as window, door etc. from a point cloud by using **OPEN3D.**
  - Find the direction of user:
    - Worked on **Real Time Data**(Scanned room).
    - Implemented a **logic** by which the system was able to find the position of camera using given random values.
    - **Optimized** Models to achieve accuracy of 99.7% and up to 10MSec time Consumption.

- **Calefyb Technology, Noida, India** ***(Feb 2021 – Dec 2021)***

  Software Engineer

  - Worked with upwork's clients:
    - Worked on real time problems.
    - Client connected and we solved their problems related to backend (language used python, sql)
  - Billing Software:
    - Designed a frontend (Wordpress) according to client **ENTAB** and prepared a pipeline for backend according to their account's formulas.
    - Designed this whole software within a deadline in which maximum time consumption of any calculation was less than 0.1 msec.
  - Face Recognition Based Attendance System:
    - **Exported Data** from camera feeds.
    - Implemented a **deep learning pipeline** for performing **face detection** using **OpenCV**, **TensorFlow, keras.**
    - Used **Raspberry Pi 4B** for taking clear pictures and reducing time complexity.
    - **Optimized** Models to achieve accuracy of 98.4%.

Education:

- Master of Technology (**M.Tech.**) in Data Analytics - **84%** ***(Jul 2019 – Jun 2021)***
  Jaypee Institute of Information Technology, India
- Bachelor of Technology (**B.Tech.**) in Information Technology - **74.9%** ***(Aug 2014 – Jun 2018)***
  Dr. APJ Abdul Kalam Technical University, India
- *Intermediate - **72%*** ***(Jul 2013 - Jun 2014)***
  S.G.I. College, Ghaziabad, India
- *High School - **82.9%*** ***(Jul 2011 - Jun 2012)***
  M.G.I. College, Agra

Personal/Freelance Projects:

- Weed And Crop Detection
  - Developed a system, in which by taking a picture, users can get to know the exact amount of crop from a field.
  - Using pandas, **Feature Engineering** (including feature selection, variable transformation, categorical encoding, feature scaling), **Model Training**, **Evaluation Metrics.**
  - **Tested** on casual images(randomly from google), achieved accuracy as 98.1%.

- Threat Detection Using Surveillance Camera
  - Developed a system, designed for women safety, if anyone would try to attack from any threat then the system would be able to give alerts to the nearest security room immediately.
  - Used data of knives and few guns and by labeling them into classes , used pandas, **Model Training(**used 1000+ images to train the data and get accurate results**)**, **Evaluation Metrics.**
  - Used **real time** data to get exact results in terms of accuracy or time consumption, achieved with an **accuracy** of **99.7%.**

- Skin Diagnosis cancer (Master's Thesis)
  - Designed a model in which the system is able to identify the spot is a casual one or a type of skin cancer.
  - Tried to improve the accuracy of skin lesions which is based on two main strategies, the first is training the models deeply on previously trained models. The second estimates the number of images within the seven categories of database used.
  - Used methods such as DenseNet, Inception, And resnet methods. And by using these methods or architecture systems got better results with high accuracy which helped humans to save their time towards their skin related serious issues.
  - Used **CNN** to classify all the classes and to get more accurate results.
  - **CNN** models verified 10000 images during the training process. Before submitting images as you enter the model, all of these images were previously processed. During the training process, the parameters of the **convolution layers** are set according to the output. And after implementing training by using the HAM10000 **dataset** we got results with 91% accuracy by applying the DenseNet method.

- Coronavirus Prediction
  - **Scraped Data** from the official website https://www.worldometers.info/coronavirus/
  - Implemented **predictive scoring** module using **supervised Algorithm** to predict the

possibility for upcoming deaths and cases.
    - **Automated** on my twitter account to aware people.

Publications:

- IJSART-International Journal for Science and Advance Research In Technology