ADITYA MISHRA

mishraaditya6991@gmail.com | +91 9029080380 | https://adimyth.github.io/portfolio

https://github.com/adimyth | https://medium.com/@adi_myth | LinkedIn

SUMMARY

I am a Machine Learning Engineer working at the intersection of Machine Learning and Software Engineering.

SKILLS

Python, ReactJS, Django, AWS, Heroku, CircleCI, Hadoop, Deep Learning Frameworks like - Keras, TensorFlow, Fast AI

EXPERIENCE

June 2018 - PRESENT

difference-engine.ai, Mumbai - Machine Learning Engineer

- ***Machine Learning for FinTech***
  - Worked on a credit-scoring system for a B2micro lending platform.
  - Involved schema transformation from product database (OLTP) to analytics-ready format (OLAP).
  - Developed a credit scoring model using tree based ensemble (sklearn random forest, **xgboost, catboost**) and deployed it as a Flask app with JSON Web Token based authentication and deployed using Waitress.
- ***Automated Report Generation Tool for a SEO Company***
  - Developed a reporting tool that processed **130GB** of data and generated more than **34, 000 reports** in just **6 hours**.
  - This was achieved by developing a **ETL pipeline** that preprocessed data from a **MySQL** database using **SQLAlchemy**, pandas and multiprocessing libraries. Implemented **CI/CD workflow** using CircleCi which ran **integration** and **unit tests** before deployment. The ETL pipeline was scheduled to run frequently using **AirFlow.**
  - It was deployed as a **react app** using **gunicorn**.
  - The system provided call to action that helped increase the number of visits, goal conversions, user engagement and decrease the bounce rate.
- ***Predicting Likelihood Of Students Dropping-off From Online Courses***
  - Given a student on any day of the online course, we developed a system to predict the drop off risk for that student based on his engagement history.
  - The data was aggregated and collected from communication methods such as emails, calendar invites, messages, calls and chat, between the Linc fellow and the students.
  - Trained a Catboost model as the data contained many categorical features and achieved 0.96 ROC AUC score.

- End-to-end model training and testing was done on **AWS** ec2.

- ***Software Engineering***
  - Developed a web app to annotate legal data which involved **scraping** supreme court reports from 1950 to 2019 using **requests and beautifulsoup,** parsing judgement texts using **tika.**
  - Saving the data on **RDS** for retrieval and updation.
  - Deployed the app on **Heroku** using **gunicorn.**
- ***Deep Learning on Identity Documents***
  - Prototyped a system to automatically detect and extract data from images.
  - This was achieved by training an **Object Detection** model (Mask R-CNN) that extracted out the document (multiple) in an image and passing them through **tesseract ocr** to extract data from it.
  - Stress test the system using **Locust**.
  - The entire pipeline was then served as a **flask** api.

October 2017 - January 2018

GreyAtom School of Data Science, Mumbai - Machine Learning Intern

- Develop data science and machine learning course materials for internal training in financial sector.
- Work with software development teams to integrate machine learning models into GreyAtom's Learning Platform.

October 2017 - November 2017

Accelo Innovation Private Limited, Mumbai - Machine Learning Intern

- Testing deep learning algorithms to detect instances of objects in real-time
- Implemented Object Distance Measurement by Stereo VISION to approximate distance between an object and a camera

ACHIEVEMENTS & CERTIFICATIONS

- ***Intel Scene Classification Challenge - AnalyticsVidhya (2019)***
  - **Top 20 rank i**n Intel Scene Classification challenge among **3000+** competitors
  - Used transfer learning with ResNet50 backbone initialised with places365 weights.
  - Used FastAI to build models, which has many state of the art techniques such as Mixup Augmentation, Cyclic LR, Ensembling, TTA.
- ***TCS Codevita Season - TCS (2017)***
  - Ranked in **top 1%** in Codevita season among **1 lakh** students.
  - The coding contest involved solving problems based on Algorithms, Data Structures, Graphs theory, etc
- ***Mumbai Hackathon - DBIT (2016)***
  - 2nd Rank at Mumbai Hackathon organized by DBIT (2016)
  - Developed a Deep Network (CNN) classifier to detect skin cancer from skin images

    - Built another Deep Network (CNN) to detect the severity of Diabetic Retinopathy
    - Transfer learning with InceptionNet using augmentation and image processing
    - The models were deployed as a web app using Flask and Materialize CSS.
- Oracle Certified Professional, Java SE 6 Programmer (2017)

EDUCATION

M.H Saboo Siddik College of Engineering, Mumbai - B.E in Computer Science (**8.8** CGPA)

August 2014 - June 2018