

# ADITYA MISHRA

mishraaditya6991@gmail.com | +91 7021590384 | [adimyth.com](http://adimyth.com)

<https://github.com/adimyth> | [https://medium.com/@adi\\_myth](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 (Django, Flask, sklearn), ReactJS, Cloud (AWS, Heroku), Deep Learning (Keras, TensorFlow, FastAI, PyTorch)

## EXPERIENCE

June 2018 - PRESENT

**difference-engine.ai, Mumbai** - Machine Learning Engineer

- **Credit-scoring for B2MicroSME Lending Platform**
  - ETL Pipeline: Schema transformation from product database (OLTP) to analytics-ready format (OLAP), orchestration and monitoring using **AirFlow**
  - Modelling: Tree-based ensembles (**sklearn** random forest, **xgboost**, **catboost**)
  - Deployment: **Flask** api with **Waitress**, JSON Web Token based authentication
- **Digital Marketing Automated Report Generation**
  - ETL pipeline
    - Processed **130GB+** data, generated **34,000+** reports in **6 hours**.
    - Stack: **Pandas+multiprocessing, MySQL+SQLAlchemy**
    - Best Practices: CI/CD using **CircleCi**, integration and unit tests using **pytest**, ETL pipeline orchestration and monitoring using **AirFlow**
    - **Frontend: ReactJSReactJS**
  - Report Generation: From 250+ metric movements, the product identified the ones with highest impact on visit counts, goal conversions, user engagement and bounce rate optimisation for reporting.
- **Student Churn Prediction from student-engagement data**
  - Feature Engineering: Features extracted from student emails, calendar invites, text messages, call recordings and chat logs on **LINCtrac** platform, between support staff and students
  - Modelling: Trained tree based models such as random forest, xgboost with **catboost** performing the best and achieving **0.96 ROC-AUC** score.
  - End-to-end model training and testing was done on **AWS** ec2.
- **Web App to annotate Supreme Court Judgments**
  - Scraping and Parsing: Supreme Court Judgments from 1950 to 2019 using **requests**, **beautifulsoup4**, and **tika-python**.
  - Deployment: **RDS**, **Heroku** and **unicorn**

- 
- **Information Extraction from Identity Documents (PAN Cards)**
    - Modelling: Object (PAN Card Image) Detection and Extraction with Mask R-CNN, Text Extraction with **tesseract ocr** (pytesseract)
    - Deployment: **Flask** api, api responsiveness stress-tested with **Locust**

October 2017 - January 2018

### **GreyAtom School of Data Science, Mumbai - Machine Learning Intern**

- Developed data science course materials for internal training in finance
- Integrated 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**

- **22nd rank among 3000+ competitors in Intel Scene Classification Challenge - AnalyticsVidhya (2019)**
  - ResNet50 backbone initialised with places365 weights (transfer learning)
  - Used FastAI to build models, which has many state of the art techniques such as Mixup Augmentation, Cyclic LR, Ensembling, TTA
- **Top 1% among 1 lakh+ competitors in TCS Codevita Season V - TCS (2017)**
- **2nd Rank among 60+ teams in Mumbai Hackathon - DBIT (2017)**
  - Detected Skin Cancer & Diabetic Retinopathy from images (Deep Learning)
  - **InceptionNet** backbone using Tensorflow (transfer learning)
  - Deployment: Web app using **Flask** and **Materialize CSS**
- Oracle Certified Professional, Java SE 6 Programmer (2017)

## **EDUCATION**

**Mumbai University**, August 2014 - June 2018 : B.E in Computer Science (**8.8** CGPA)