Kuldeep **Gautam**

kgautam01.github.io | cs20mtech01004@iith.ac.in | kuldeep.gautam075@gmail.com

LINKS

Github:// **kgautam01**
LinkedIn:// **kgautam01**
Leetcode:// **kgautam01**
Hackerrank:// **kgautam01**

COURSEWORK

GRADUATE
Advanced DSA
Foundations of Machine Learning
Deep Learning for Vision
Compiler Optimizations
Advanced Compiler Optimizations

UNDERGRADUATE
Data Structures and Algorithms
Operating Systems
Database Management Systems
Compilers

ONLINE
DL Specialization, **deeplearning.ai**
Machine Learning A-Z, **Udemy**
DSA using Python, **IITM-NPTEL**

SKILLS

PROGRAMMING
Python • C • C++
HTML/CSS • LaTeX • Markdown
Portable Bash Script • Bash

TECH STACK
Machine Learning
Deep Learning
Computer Vision
Natural Language Processing

LIBRARIES
Pytorch • Tensorflow • Keras
Scikit-learn • NLTK • OpenCV
Numpy • Pandas

EDUCATION

INDIAN INSTITUTE OF TECHNOLOGY, HYD **JAN '20 - DEC '22**
M.Tech (Research) in Computer Science & Eng.
GPA: 9.0 / 10.0

AJAY KUMAR GARG ENGINEERING COLLEGE, GZB **JULY '15 - JUNE '19**
B.Tech (Hons) in Computer Science & Eng.
GPA: 7.9 / 10.0

EXPERIENCE

INDIAN INSTITUTE OF TECHNOLOGY, HYD | Research Asst. **JAN '20 - DEC '22**
- Working as a Research Asst. under the supervision of **Dr. Ramakrishna Upadrasta** at Scalable Compilers for Heterogeneous Architectures Group, IIT Hyderabad.
- Actively carrying out research based on Machine Learning in Compilers and Programming Languages.
- Paper presentations and creating surveys for research papers for state-of-the-art works to understand the approach and reproducing & evaluating the results.
- Mentoring juniors in the research group.
- Teaching Asst. for Compilers-2, and Compiler Optimizations.
- Responsible for maintenance and updates of the team website.
- Organised ***ACM India Summer School, 2021*** on Programming Language Analysis and Optimizations, hosted by IIT Hyderabad, in collaboration with ***ACM India Council*** and ***NVIDIA***.

SAMSUNG RESEARCH INSTITUTE, NOIDA | Software Eng. **JULY '19 - DEC '19**
- Part of Project Leading group, handling project management, code quality & compatibility of development, OS-upgrades, and maintenance release projects.

STRATEGIC IP INFORMATION (SIPI IP), NOIDA | DS Intern **JUNE '18 – JULY '18**
- Implemented a computer vision model for image similarity.
- Implemented a python module for translating large data in foreign languages to English for better text processing & creating word clouds.
- Created a chrome extension for a marketplace for scraping the data.

ACADEMIC RESEARCH

CODE VULNERABILITY DETECTION **NOV '21 – PRESENT**
- Working with **S. VenkataKeerthy** under the supervision of **Dr. Ramakrishna Upadrasta** to generate architecture agnostic distributed representation of binaries and using it to detect vulnerability in the binaries.
  | ***python, c, bash, angr, valgrind***

CODE SUMMARIZATION & RETRIEVAL ENGINE **FEB '21 – PRESENT**
- Working with **S. VenkataKeerthy** under the supervision of **Dr. Ramakrishna Upadrasta** to develop a language agnostic code summarization and retrieval engine for generating code summaries, performing code search and code reconstruction, providing language agnostic embeddings for all tasks in future.
  | ***python, bash, pytorch***

COFO **NOV '20 – FEB '21**
- Worked with **S. VenkataKeerthy** under the supervision of **Dr. Ramakrishna Upadrasta** to create a dataset named COFO using a python-selenium-beautiful soup based Codeforces scraper consisting of ***809 unique problems, 369K source codes in C, C++ (v11, v14, v17), Python3, and Java (v8, v11)*** suitable for code classification and code tagging task.
  | ***python, bash, selenium, beautiful soup***

ACADEMIC PROJECTS

ML/DL PROJECTS

ScratchML: **AUG '21 - DEC '21**

Collection of ML algorithms implemented from scratch.

- ***Logistic Regression:*** Numpy based implementation of logistic regression with log loss as the loss function. It includes utility for fit, predict and score functions, similar to scikit-learn API.
- ***Decision Trees:*** Implementation of entropy and gini-index based decision tree classifier. It also includes implementation of K-Fold cross validation from scratch.
- ***Random Forests:*** Implementation for entropy and gini-index based random forest classifier, making use of the decision tree module built in other project. It also provides utility functions for fitting, predictions, out of the bag score and performs at par with scikit-learn RFC API in accuracy. Sensitivity to max_features parameter and OOB error has also been explored.
- ***Feed Forward Neural Networks:*** Tensorflow based implementation of FF-NN. Adam optimizer has been used to perform the backward pass only using TF API. Results in an accuracy of 96.6% on MNIST dataset.
- ***Density-Based Spatial Clustering of Applications with Noise (DBSCAN):*** Implementation of DBSCAN method of clustering. Used euclidean distance for getting points in epsilon neighborhood and separates the points into core, border and noise points. Only core points and border points are used to define boundary lines for clusters. Performs at par with scikit learn implementaton.

| ***python, numpy, tensorflow***

Basic: **JAN '18 - MAY '18**

Collection of basic Deep Learning projects.

- ***DNN based Face Detection:*** ResNet10 architecture based face detection project implemented using OpenCV and Caffe implementation of ResNet10 architecture(pretrained).
- ***Automated Essay Scorer:*** Word2Vec based essay scoring project using LSTMs to process the word vectors in the sentence. Pipeline includes: removal of stopwords, punctuations, special characters, lowering of cases, removing numeric values, stripping spaces, and tokenization. K-Fold CV for determining the results on validation dataset and ensuring performance on test dataset. The results are then evaluated for Mean Squared Error, Variance and Cohen's Kappa Score.

| ***python, pytorch, nltk, gensim, opencv, scikit-learn***

Kaggle: **MAR '18 - MAY '18, NOV '21 - DEC '21**

Collection of classification/regression challenges on Kaggle.

- ***The Titanic competition (Accuracy: 78.4%)***
- ***Driver Fault competition (Accuracy: 84.38%)***
- ***The Housing competition***
- ***Taxi Fare competition***

Pipeline includes:

- Preprocessing of data.
  - Dealing with missing values using impuation methods like mean, median, etc.
  - Coversion of categorical values using label encoder followed by conversion to one-hot vector for training.
  - Feature engineering for merging features and create a new better feature like age bands, days of the week, month of the year, etc.
  - Feature reduction methods like forward selection, backward elimination and PCA for dimensionality reduction.
- Model Building, Hyperparameter-Tuning using grid search and Evaluation using eval metrics such as confusion matrices, accuracy, etc.

| ***python, scikit-learn***

METHOD NAME PRE-PROCESSOR **NOV '20**

Implemented a text preprocessor using trie data structure to preprocesses short method names to meaningful method names as given in English dictionary alongwith a custom dictionary containing CS related words/acronyms.

| ***python, nltk***

ACHIEVEMENTS

- 96.50 percentile in GATE 2019
- Organizer for the ACM India Summer School, 2021 on Programming Language Analysis and Optimizations, hosted by IITH.
- Global Rank: 2001/10177 in Google Hash Code 2022 Qualification Round.