Abhishek Mukherjee
567-806-0840
abhi0787@gmail.com
LinkedIn: tinyurl.com/abhilnkdin0787
GitHub: github.com/amukher3
Tableau:https://public.tableau.com/profile/abhishek.mukherjee#!/

ABHISHEK MUKHERJEE

SUMMARY

- A highly motivated, dynamic and hardworking professional with 3+ years of experience in Machine Learning, Deep Learning and Data Science.
- High problem-solving abilities with very strong skills in Python.
- Knowledge and skill in implementing Data Engineering pipelines.
- Strong fundamentals in Data Structures and Algorithms.
- Great communication skills.

EDUCATION

UNIVERSITY OF TOLEDO August 2016-August 2019
MASTERS IN COMPUTER SCIENCE

INDIAN INSTITUTE OF TECHNOLOGY, DHANBAD July 2014-July 2016

COMPUTER SKILLS

Languages:	Python, C, MATLAB
Web:	Django, Bootstrap, CSS3, jQuery, JavaScript, HTML5,Flask
Databases:	MySQL, Oracle, SQL,MongoDB,Neo4j
Data Engineering	PySpark, Hive
Data Science	Pandas, Sklearn, Spacy, Gensim, Scipy, Numpy, TensorFlow, Keras, PyTorch, NLTK, OpenCV Beautiful Soup, Seaborn, Matplotlib, Plotly, Cufflinks
Data Visualization	Tableau, PowerBI
Platforms:	Windows, Linux

EXPERIENCE

January 2020 to Present — **Zillionix**, Freemont, California
Title: (Data Analyst Intern)

Responsibilities:

- Web-scrapped craigslist ads and further clean, process and parse them to extract useful information.
- Designed and implemented Data Engineering pipelines using Spark and Hive to further process the parsed data.
- Designed and reported results using Tableau, Plotly.
- Web-scrapping was done using Beautiful Soup further cleaning and processing was done using Pandas and Regular Expressions.

Technologies: Python, Regular Expression, Pandas, Beautiful Soup, Tableau , Plotly, Windows, Linux.

August 2016 to August 2019 — **UNIVERSITY OF TOLEDO,** Toledo, Ohio
Title *Research Assistant/Data Scientist*

Responsibilities:

- Applied Statistical Learning techniques and Signal Processing theory to develop statistical modeling methods for Physiological data.
- Generative modeling of severely under-sampled electrodermal activity data to detect and reconstruct major skin conductance activity changes.

Technologies: MATLAB, Windows, Linux.

PROJECTS:

- **Sarcasm vs. Satire:** *An effort to separate sarcasm from satire in English language sentences using deep learning methods. A Bi-directional LSTM model was used to separate Sarcasm and Satire in English language sentences.*
  Technologies used: *Python, Keras, Sklearn, Pandas*
  Repository: github.com/amukher3/Sarcasm_vs_Satire
- **Idiom Modeling:** *Modeling of English language idioms from unstructured data using Latent Dirichlet's Allocation (LDA) approach to Topic modeling. Initially the data was scrapped off from various web-sites using Beautiful soup and then processed and cleaned using Pandas after* which the modeling was done using NLTK and SPACY.
  Technologies used*: Python, Beautiful Soup, Pandas, NLTK, Spacy, Sklearn.*
  Repository: github.com/amukher3/Idiom-modeling
- **Bay Area Housing:** *Exploratory data analysis of unstructured data obtained from Craigslist ads which were posted for potential tenants. Data was scrapped off using Beautiful Soup after which analysis was done using Pandas.*
  Technologies used**:** *Python, Beautiful Soup, Pandas*
  Repository: github.com/amukher3/Bay-area-housing
- **Frontal face controlled cursor***: A novel and extremely simple way to control the mouse wirelessly, just with face movements is presented here.*
  Technologies used**:** *Python, OpenCV, Haar Cascade*
  Repository: https://github.com/amukher3/Frontal-face-controlled-mouse
- **Respiratory Sinus Arrhythmia Estimation:** *An effort to separate arrhythmic component from the Heart Rate time series structure, using Knowledge-based dictionaries.*
  Technologies used: *MATLAB*
  Repository: github.com/amukher3/Respiratory-Sinus-Arrhythmia-estimation
- **Plasma cell identifier:** *A method to identify normal plasma cells from infected blast cells using 2D-CNN over optical microscope bone marrow images. Model can be currently deployed as a web based framework locally using Flask/Python.*
  Technologies used*: Python, Flask, Keras, Numpy*
  Repository: github.com/amukher3/Plasma-cell-identifier
- **DCGANs X-ray images**: *Using DCGANs to generate X-ray images of Sacroiliac joints infected by* Ankylosing Spondylitis.
  Technologies used: *Python, PyTorch*
  Repository: https://github.com/amukher3/DCGANs_X_ray_Images
- **Exploratory Data Analysis:** *A repository containing detailed exploratory data analysis of various data-sets found in Kaggle and other websites.*
  Technologies used**:** *Python, Pandas, Numpy, Sklearn*
  Repository: github.com/amukher3/Exploratory_Data_Analysis

CERTIFICATIONS:

- Machine Learning A-Z: Hands on Python & R in Data Science.
- Neural Networks and Deep Learning.
- Improving Deep Neural Networks: Hyper-parameter Tuning, Regularization & Optimization.
- Sequence Models.
- Deploying machine learning models with Flask for beginners.
- Natural Language Processing with Python.
- Power BI A-Z: Hands-on Power BI training for Data Science
- Tableau 10 A-Z: Hands-on Tableau Training for Data Science
- Deep Learning and Computer Vision A-Z: Open CV,SSD& GANs.
- Taming Big Data with Apache Spark and Python.