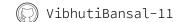
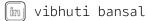
# Vibhuti Bansal

# Data Scientist / ML engineer









m vibhuti bansal bansal.vibhuti25@gmail.com

#### INTERNSHIPS

# INNOMATICS RESEARCH LABS | DATA SCIENCE INTERN

April 2021 - July 2021

- → Understood and implemented concepts of Statistics and made Music Recommender Systems using Python
- → Used QQ plots. Box Cox Transformation and IQR to understand the data and Conducted p-test, z-test and hypothesis testing on it.

# ROBOIOTICS SERVICES LLP | ML INTERN FLUTTER DEVELOPER

**August 2020 – October 2020** 

→ Optimized Deep Learning Model for offline deployment with App Compatibility using tflite. Made app UI from scratch with FireBase as backend.

ML INSTRUCTOR | WAC - WHAT AFTER COLLEGE

#### RESEARCH PUBLICATIONS

#### 4-CLASS LUNG DISEASE DETECTION INCLUDING COVID | PYTHON

Accepted for 26th International Conference of Pattern Recognition

- → Preprocessing technique for contrast enhancement along with correlation based metric undersampling
- → Derived of a new custom loss function to reduce class imbalance.
- → Training of various SOTA machine learning models on the derived dataset and obtaining 4 class classification F1 score 96 percent.

# ABUSE DETECTION IN 13 LOW RESOURCE INDIC LANGUAGES | PYTHON

Submitted: Transactions on Asian and Low-Resource Language Information Processing

- → Cleaning of dataset including transliteration, translation of dataset in order to extract maximum performance.
- → Highlighting the importance of emoji in hate speech detection. Training transformer based models to obtain maximum F1 score.

#### ENHANCED DCNN FOR CLASSIFYING 8 ICD FORMS | LINK

Published: Applied Sciences as part of the Special Issue Human-Computer Interaction for **Industrial Applications** 

→ Model to classify images into 8 Indian Classical Dance Forms. It uses Image Truncated Thresholding and ResNet50 as base model.

#### **PROJECTS**

#### **GEOSIS** [LINK TO PROJECT]

- → Dataset scraped from govt. sources for a time period of 20 years. Models were built on this data for future prediction of the climate cover.
- → Visual analysis dashboard from google earth engine stats was utilised to observe the NDVI change as well as climate change over time in different regions of India.

### COVID - 19 VACCINE PRIORITIZER USING CLUSTERING [GITHUB]

→ Data Preprocessing, Feature selection, and undupervised learning (KMeans) were used in a pieline. SHP data was used to visualize the results on India's map using GeoPandas.

#### SKILLS

#### **PROGRAMMING**

C++ • C • Python • Dart

#### LIBRARIES/FRAMEWORKS

Tensorflow • Keras • Scikit Learn • OpenCV • Numpy • Pandas • Matplotlib • GeoPandas • FireBase • Flutter • SQL

#### **TOOLS/PLATFORMS**

Git • GCP • Juptyer Notebook • VSCode • Android Studio • MySQL • Advanced Excel

# **EDUCATION**

# BHARATI VIDYAPEETH'S **COLLEGE OF ENGINEERING**

BACHELOR OF TECHNOLOGY (CSE) 2019 - 2023 CGPA: 9.22 / 10.0

# SETH ANANDRAM JAIPURIA SCHOOL

Class XII: 95.2 % Class X: 10 CGPA

#### PARTICIPATION

WIE HACK 3.0

1st Runner-Up WIEhack 3.0

#### **EY TECHATHON 2**

Selected in top 6 teams

#### CAMPUS BLOCKS BVP

**Vice Chairperson**