

# Abish Pius


Data Scientist



## Personal details

 abishpius@gmail.com

 +1 (412) 482-2088

 abishpius.itsben.in

 linkedin.com/in/abish-pius-424566b3

## Skills

- Python
- R
- Bash
- SQL
- PySpark
- TensorFlow
- PyTorch
- OpenCV
- R Studio
- AWS
- Azure
- OCI
- Kedro
- VS Code
- Git
- Docker

## Work Experience

### Data Scientist

Jan 2021 - Present

United States Steel Corporation, Pittsburgh, PA

- Designed iron ore price 6-month forecasting model alongside a hedging decision strategy with expected returns >10% compared to baseline. Productionalized ingestion and modelling pipelines using kedro.
- Built statsmodel for 1 to 3 month grinding materials price forecast. Led discussion with business team on sourcing new data, pitfalls of relying on limited data, and preventative maintenance of models.
- Counted pipe inventory for different production plants using Azure custom vision and Oracle vision tool with 98% precision.

### Freelance Big Data Solutions Expert

Oct 2021 - Present

UpWork, ID: freelancers/01d2db74395151722c

- Completed 30+ jobs, received Top Rated status and earned membership in the Stanford Talent Cloud for successful analysis on Sensomics startup data within GCP Vertex AI.
- Consulted as a Data Science Lead for Trajum Analytics improving medical appointment booking systems through a hosted model endpoint on GCP AI Platform API.
- Consulted Vitro Technology Corporation in building a POC dashboard for SNMP Thingsboard data in AWS utilizing S3, Redshift and QuickSight.

### Data Scientist

Jul 2017 - Sep 2020

McGowan Institute of Regenerative Medicine, South Side, PA

- Pre-processed large genomics data parallelly on multiple compute nodes using bash and slurm scripting. Cleaned and visualized gene sequencing data using R Bioconductor and Python Scanpy to display differences in skeletal muscle health.
- Utilized Cell Profiler CNNs to segment cells from fluorescent microscopic images and used random forest algorithms to classify young versus old muscle stem cells as well as PCA to identify distinct cell subpopulations.

## Projects

### Data Science Subject Matter Expert

Nov 2021 - May 2022

Emeritus Course Consult

- Consulted as SME for two courses, CMU's Introduction to Computer Vision and NYUT's Using Data Science to Manage Cyber Risk
- Reviewed course materials for accuracy and designed/updated questions and exercises within Canvas and Codio.

### Streamlit Web Application

Sep 2020

Deep Learning Financial Time Series Model

- Used TensorFlow to create a DL LSTM model to forecast univariate time series to specified period.
- Created web application with Streamlit and hosted project as container on Heroku.

### Machine Learning Competition

Mar 2020

M5 Walmart Sales Forecasting Competition

- Predicted Walmart quarterly sales per retail location given five-year sales data using random forest regression in Python
- Used clever feature engineering to account for inflation as well as one hot encoding store location by county
- Placed in top 30% of submitted models

## Certifications

### AWS Certified Cloud Practitioner

Jul 2022

AWS

### Data Engineering with Google Cloud Professional Certificate

May 2020

Coursera

## Education

### Masters of Science in Computational Biomedicine

2019 - 2020

University of Pittsburgh, Pittsburgh

### Bachelor's of Science in Neuroscience

2013 - 2017

University of Pittsburgh, Pittsburgh

Minors: Chemistry, Mathematics