Aritra **Biswas**
Data Scientist

linkedin.com/in/pandalearnstocode github.com/pandalearnstocode
+91-9051237918 @ pandalearnstocode@gmail.com
186, 18th B main, HAL 2nd Stage, Indiranagar, Bangalore 560038, Karnakata, India
https://www.pandalearnstocode.com/

Data scientist with experience in Python, R, C/C++, TensorFlow, Spark, Excel, SPSS, DevOps , CI/CD to manage large scale distributed applications on cloud. Passionate about reproducible data analysis and product development, mainly focused on statistical model building, distributed computing, profiling and visualization, with a solid background in statistics along with extensive experience using data insights to drive business growth. Also interested in Natural Language Processing, Text mining, Computer vision and Robotics.

Compétences

Programming Language	**Python** (Numpy, Pandas, Dask, Numba, Scipy, Scikit-learn, Tensorflow, PySpark,snakeviz), **R** (tidyverse, shiny, caret, profievis, rmarkdown), **C, Julia**, LaTeX, markdown, HTML, GW-BASIC
Spreadsheet Software	IBM SPSS, Minitab, **Microsoft Excel**
Other tools	**Docker, Git, Postman**, SVN, JupyterLab, Jupyter Notebook, **Open CV**, NP++
Operating System	**Cent OS**, Ubuntu, **REHL, Windows Server**
Web Technology	WordPress, Joomla, Hugo, Drupal
Technical skill	Classification, Regression Analysis, Support Vector Machine, Natural Language Processing, Random Forest, Text Mining, Hypothesis Testing, Dimensionality Reduction, Principal Component Analysis, Simulation, Non-Linear Constrained Optimization
Package Development	YoutubeR, MarketingPlanneR

Expérience Professionelle

Nielsen
May,17 - Present

MROI | Product Development and Innovation, AAC, Bangalore, India

- Development of Marketing Mix Modeling and Market Planner tool using Python based on Hierarchical Linear Model (HLM), Generalized reduced gradient decent (GRG2) and Monte Carlo Simulation.
- Profiling and Optimizing code using SnakeViz, profvis, Rcpp, Numba, Cython and Dask.
- In depth understanding of additive and multiplicative model in Marketing Research Domain.
- Monitoring and maintenance of staging and production server.
- Extensively used Non-Linear box constraint based optimization (gradient and non-gradient based methods) with help of various python and R libraries.
- Vendor management and continuous communication with user community.
- Developing methodology related documents for internal and external communication.

Python | Flask | Numpy | Pandas | Postman | Jenkins | R | tidyverse | docker | CentOS

Chegg, Inc.
Jan, 16 - May, 17

Subject matter expert, Statistics, Data Science

- Teaching graduate and under-graduate students statistics, mathematics and economics.
- Helping with academic assignments in MLA, APA and other academic formats.
- Helping students writing research paper and proposal for journal and conferences.
- Developing reading materials, solving text book based problems and various concepts related to statistics and computer science with graphical illustrations.
- Designed multiple online tests, assignments and programming assignment tests using shiny server pro including user authentication.

Tutorial | Academic Writing | Research Methodology | Statistics Tutor

Blue Copper
Jnue - August, 16

Business Analyst (Intern) , Blue Copper Technologies, Calcutta, India

- Developing predictive models, information extraction and visualization from web using interactive dashboard solution.
- Predictive model building (linear regression, logistic regression) and residual diagnostics using graphical methods. Extensively used ggplot2 and shiny to build interactive dash-boarding solution.
- Basic text processing, cleaning, stemming, corpus formation, association mining, term document matrix and TF-IDF construction in R.
- Used Named Entity Recognition using Max Entropy classifier. Developed Latent Dirichlet Allocation and Correlated Topic Model based topic models.
- Applied shallow neural network based models such as Word2Vec and Doc2Vec, N-grams, Skip-grams, tidytext and lexicon based methodologies for sentiment and polarity analysis using R. Extensively used NLP, OpenNLP, tm, tidytext and syuzhet package for textual analysis.

R | htmlwidget | Word2Vec | Doc2Vec | shiny | Regression | NLP | Web Scraping | NER | MaxEnt | Sentiment Analysis

Academic and Individual Projects

Facebook Bot
February,2018

Implemented a Facebook chatbot, (using dialogue flow , facbook API in python)

- Created an initial Facebook chat-bot for automated reply and small talk.
- Connected it with web-hook and personal blog. Used dialog flow for data storage and automated replies using deep learning based models trained on user communication data.
- **Next Phase :** To connect it with the Google assistant and voice API.
- **Further Development :** To test this bot in a scalable platform for intent and behavioural analysis. Here is a short video easily of the chat bot in the initial stage.

Python | Facebook API | Chatbot | dialogflow | WebHook | Flask | ngrok

Computer Vision
Jan,2018

Object detection and classification using Tensorflow, Darknet , YOLO V2, and Tiny YOLO

- Implemented Tensorflow and Open CV for image classification.
- Integration of GPU for video processing and object classification using pre-defined models.
- Classified object in a live video using web camera and pre-trained YOLO V2 and tiny YOLO.
- Training new models for classifying new objects using web scrapping with higher level of accuracy. Here is a smaller video demonstrating classification using live cam feed, tensorflow and opencv.

Python | OpenCV | tensorflow | raspberry pi | YOLO | GPU

YouTubeR

July,2015

A comprehensive package to access and analyse data from YouTube Data API Version 3 via R, (Hosted in Bit bucket), (Incomplete Documentation)

- This package converts the complex data structures, sparse data, nested lists of the from JSON to class S3 object, does missing value treatment, graphical analysis and sentiment analysis.
- Application of rvest, jsonlite, httr, RCurl (for web mining and scarping), tidyr, dplyr,purrr, broom,lubridate, strinigi, stringr (for data manipulation) and Packrat (for reputability and dependency management and robust development).
- Here is the link for applications developed using this package, 1) Indian News Channel Polarity Analysis, 2) Geospecial viewership monitor of Youtube Channel Being Indian, 3) Task scheduling in R.

rvest | jsonlite | RCurl | tidyverse | purrr | broom | packrat | Youtube Data API V3 | G+ API | Facebook API

Dissertation

June, 2016

Server based interactive programming in R to visualize Maximum likelihood estimate, Supervisor : Dr. Kapil Kumar, Assistant Professor

Delhi University Department of Statistics

Simulating data for a given user specified probability distribution (i.e. family or distribution name and parameter space). This interactive application uses Shiny Server to visualize the log-likelihood of the simulated data in a contour plot and obtains MLE estimates of the population parameters by solving a system of non linear equations easily

Here is the link for other academic projects, 1) Visualizating MLE, 2) SPSS, 4) C, 5) Minitab, 6) R

shiny | MLE | Mutivariate Data Visualization | SPSS | C | Minitab | R | Excel

Programming Languages

R	●●●●●
Python	●●●●○
C/C++	●●●●○
Julia	●●●○○

Interests

- Cooking
- Blogging
- Reading
- Robotics

Educational Qualification

M.Sc. (Statistics)	University of Delhi, Delhi, India.
B.Sc. (Statistics)	Presidency College, University of Calcutta, Kolkata, India

Online courses and certifications

Stanford Online	Statistical Learning, 2016
IIMB, edX	Predictive Analytics, 2015

Références

Jon Snow
Lord Commander, Night's Watch
@ john.snow@nightwatch.org
+687 987 654

Eddard Stark
King of the North, Winterfell
@ e.stark@winterfell.org
+33 1 23 45 67 90

Eddard Stark
King of the North, Winterfell
@ e.stark@winterfell.org
+33 1 23 45 67 90