Chamberlain Mbah

Curriculum Vitae

"Anyone who has ever made anything of importance was disciplined. Discipline is key."

Work Experience

2019– —— **Data/Research Scientist**, *Yields.io*, Brussels, Belgium.

- Build generative models (Generative adversarial Networks GANs, variational autoencoders (VAEs) for synthetic data generation. Used synthetic data for fraud detection and data privacy projects
- Played a leading role in project that used Bidirectional Encoder Representations from Transformers (BERT) to perform sentiment analysis.
- Used Bidirectional Encoder Representations from Transformers to predict regime change in stock prices.
- Hands on experience on designing and implementing extensive IR/FX derivatives pricing models and user interfaces for front office usage.
- Manage various big datasets pipelines, leveraging power of PySpark on distributed file systems such as Hadoop.
- Apply various statistical test to check data quality.
- Apply state of the art techniques like autoencoders for anomaly detection and outlier detection.
- Making sure models in production for our clients adhere to the Assessment List for Trustworthy Artificial Intelligence (ALTAI). ALTAI recommends that models are explainable, fair and unbiased.
- Create Machine learning (ML) pipelines from data versioning with hudi to serving models with kubeflow and solving complex business problems.
- Containerise ML models (docker and docker compose).

2019– —— **Visiting lecturer**, *African institute of Mathematical Science*, Mbour, Senegal.

- Main lecturer of the course High Dimensional Data Analysis
- Co-lecturer of a big data course where PySpark, big query, SQL are introduced to the students
- Main lecturer of the course Convolutional Neural Networks
- Supervisor of Master's thesis
- Research grants applications

2018–2019 **Lead data scientist**, *Tobania*, Brussels, Belgium.

- Monitored and scheduled data scientists on different project.
- Brainstorm on the approach to take in tackling different ML problems.
- Built model and attrition models, serve this model with REST API endpoints.
- Designed multiple experiments for testing marketing approaches, analysed the results with simple statistical test and display findings to the C-level.

2017–2018 **Research Assistant**, *Gent University*, Gent, Belgium.

2013–2017 **Ph.D. research**, *Gent University*, REQUITE, FP7 project.

Awards

2013 FP7 Scholarship for Ph.D. research under the REQUITE Project

Programming skills

- Advanced knowledge: python (tensorflow,pyspark)–R (tydiverse)–SQL–git – docker–linux – cloud (AWS,GCP,Azure)
- Working knowledge: hadoop–Hive–Spark

Education

2013– 2017 **Ph.D. in Statistical Data Analysis**, *Gent University*, Gent, Belgium.

2011– 2013 **Masters in Statistical Data Analysis**, *Gent University*, Gent, Belgium.

2008– 2010 **Masters in Studies in Mathematics**, *University of Buea*, Buea, Cameroon.

2005– 2008 **Bachelor in Mathematics**, *University of Buea*, Buea, Cameroon.

Selected publications

- **C Mbah**, J De Neve, O Thas. *High dimensional prediction of binary outcome in the presence of between-study heterogeneity.* Statistical Methods in Medical Research (in press). -

- **C Mbah**, K De Ruyck, S De Schrijver, C De Sutter, K Schiettecatte, C Monten, L Paelinck, W De Neve, H Thierens, C West, G Amorim, O Thas and L Veldeman (2018). *A new approach for modeling patient overall radiosensitivity and predicting multiple toxicity endpoints for breast cancer patients. Acta Oncologica, 57:5, 604-612, DOI: 10.1080/0284186X.2017.1417633* -

- **C Mbah**, H Thierens, O Thas, J De Neve, J Chang-Claude, P Seibold, A Botma, C West. *Pitfalls in Prediction Modeling for Normal Tissue Toxicity in Radiation Therapy: An Illustration With the Individual Radiation Sensitivity and Mammary Carcinoma Risk Factor Investigation Cohorts. Int J Radiat Oncol Biol Phys. 2016 Aug 1;95(5):1466-1476. doi: 10.1016/j.ijrobp.2016.03.034. Epub 2016 Apr 1.* -

- **C Mbah**, K Peremans, S Van Aelst, DF Benoit. *Robust Bayesian seemingly unrelated regression model. Computational Statistics . 2019.*

Interests

- Football
- Running
- Walking