Aboubacar HEMA

Liberté 1
Dakar-Senegal
✆ *+221 77 403 89 55*
✉ *aboubacarhema94@gmail.com*

Profile

Data scientist with experience in data analytics, data mining, data processing, and storytelling. Skilled in Python, SQL, Tableau, and machine learning algorithms. Successfully completed multiple data science projects, solving real-world data-driven problems. Seeking a challenging data science role to apply and enhance my knowledge and skills.

Experience

November 2022 - Present

Research Analyst, International Food Policy Research Institute (IFPRI), Regional Office for West and central Africa
- Conducted data collection and cleaning of primary data in Senegal and other (West and Central) African countries, resulting in a dataset of over 5,000 observations. - Performed economic modeling and econometric analysis in STATA, resulting in a 15- Contributed to the preparation and publication of discussion papers, reports, technical notes, and journal articles, resulting in over 10 publications in peer-reviewed journals.

April 2022 - October 2022

Data Analyst, International Consultant, International Organization for Migration (IOM) , Senegal
- Conducted analysis and implementation of Mental Health and Psychosocial Support (MHPSS) surveys (round 1-3) and endline surveys for Senegal, Guinea, Nigeria, and The Gambia as a data analyst with the International Organization for Migration (IOM). - Leveraged data processing and analysis skills to collect, clean, and interpret survey data to generate insights and recommendations for program improvement. - Successfully collaborated with cross-functional teams to prepare and present reports on survey findings, achieving a 20% increase in stakeholder engagement and satisfaction.

May 2021 - July 2022

Research Assistant in Statistics, Senegalese Institute for Agricultural Research., Senegal
- Conducted end-to-end data analysis for multiple impact evaluations as a Research Assistant in Statistics at the Senegalese Institute for Agricultural Research, overseeing the collection, processing, and analysis of large datasets, resulting in a 25% reduction in data processing time. - Developed data quality monitoring protocols and provided guidance for enumerator training to improve data accuracy and completeness. - Coordinated with field teams to identify and address data quality issues, resulting in a 15% increase in overall data accuracy.

October 2020 - May 2021

Statistician, Consortium for Economic and Social Research , Senegal
Responsible for data collection, managing, processing and analyzing data from impact evaluations. Tasks also involved supporting enumerator training, developing protocols for data quality monitoring, and communicating with field teams on data quality issues.

March 2020 - October 2020

R package developer/Maintainer, Agricultural Research for development (CIRAD) , Senegal
- sdmApp : A user-friendly application for species distribution modelling.
  - CRAN: https://cran.r-project.org/web/packages/sdmApp/index.html
  - Github: https://github.com/Abson-dev/sdmApp

Machine Learning Projects

Automatic Review Analyzer

• Summary: Designing a sentiment analysis classifier for product reviews. • Work/Impact: Trained a classifier on a dataset of Amazon customer reviews for food products. • Performance Metric: Achieved an accuracy of 85% on the test set. • Tools/Skills: Natural Language Processing, Classification Algorithms, Feature Engineering.

Digit recognition

- Summary: Implementing various Machine Learning algorithms for classifying MNIST digits.
- Work/Impact: Used Linear and Logistic Regression, Regularization, Non-linear features, and Kernel Tricks to improve model performance. Implemented Neural Network for classification.
- Performance Metric: Achieved accuracy of 98% on MNIST dataset. - Tools/Skills: Machine Learning algorithms, Neural Networks, Python, Scikit-Learn, Tensorflow.

Collaborative Filtering via Gaussian Mixtures

• Summary: Built a mixture model for collaborative filtering to predict remaining movie ratings.
• Work/Impact: Developed Gaussian mixture model and applied it on the partially filled data matrix. • Performance Metric: Achieved an accuracy of 90% on predicting remaining movie ratings.
• Tools/Skills: Gaussian mixture model, data imputation techniques, collaborative filtering.

Text-Based Game

• Summary: Developed control policies for text-based games using reinforcement learning. • Work/Impact: Trained RL models to interact with virtual world through text using natural language commands. • Performance Metric: Achieved a success rate of 75% on complex text-based game scenarios. • Tools/Skills: Reinforcement Learning, Natural Language Processing, Text Generation.

Effectiveness of an experimental design

• Summary: Analyzed experimental design elements such as placebo effect and double-blindness, conducted hypothesis testing, and identified potential biases. • Work/Impact: Conducted multiple hypothesis testing, ran statistical models for data, and defined suitable test statistics based on research questions. • Performance Metric: Applied Holm-Bonferroni and Benjamini-Hochberg corrections and validated results using other data sets.

Genomics and High Dimensional Data Analysis

In this project, we analyze a single-cell RNA-seq data set, with the goal of unveiling hierarchical structure and discovering important genes. The data sets provided are all different subsets of a larger single-cell RNA-seq data set, compiled by the Allen Institute. This data contains cells from the mouse neocortex, a region in the brain which governs higher-level functions such as perception and cognition.

Networks Analysis

In this project, we analyze two networks data sets, with the goal of understanding the behaviors and relationships between the actors in these data sets.

Time Series Analysis

The goal of the problem is to fit the data and understand its variations.

Environmental Data and Gaussian Processes

Recommendation Systems

Breast Cancer Prediction Project

Education

February 2023 - June 2023	**Data Science and Machine Learning: Making Data-Driven Decisions**, MIT Institute for Data, Systems, and Society (IDSS),
January 2022 - January 2023	**MITx MicroMasters program in Statistics and Data Science**, – Massachusetts Institute of Technology,

Master the foundations of data science, statistics, and machine learning. Analyze big data and make data-driven predictions through probabilistic modeling and statistical inference; identify and deploy appropriate modeling and methodologies in order to extract meaningful information for decision making.

2016–2020	**Statistician Engineer Diploma**, National School of Statistics and Economic Analysis (ENSAE-Dakar), Dakar-Senegal,

Real Analysis, Probability and Statistics, Optimization, Microeconomics (I and II), Macroeconomics, Econometrics, Microeconometrics, Spatial Economics, Data mining, Big Data, Advanced database, Geographic Information Systems, Monitoring and evaluation of projects

2012-2016	**Computer Science**, *Option : Information Systems and Networks(ISN)*, Université Joseph KI- ZERBO, Burkina Faso,

Algorithms, Computer Architecture, Operating Systems, Dynamic Data Structures, Networks and Telecommunications, Web Technologies, Analysis and Design of Information Systems, Relational Databases, Database-Oriented Development Tools, Data Mining.

Skills

Statistics and programming:	R, Stata, Python, SQL
Data visualization:	R Shiny, ggplot, R leaflet, plotly, Tableau
Other tools:	Git/GitHub, LaTeX, ArcGIS, HTML/CSS, ODK
Languages:	English (proficient), French (native)

Working Papers

Ninon Sirdey, Tomoé Bourdier, Patricio Mendez Del Villar, Aboubacar Hema, Ibrahima Dedhiou, Djiby Dia. Juin 2021 – Mars 2022. Etat des lieux de la riziculture pluviale en Casamance. 26p.

Diakhaté PB, Thierry FERRE, Djiby DIA, Aboubacar HEMA. 2022. Caractérisation des PME de transformation de l'arachide au Sénégal et analyse des pratiques d'innovation. 46p.

Mamadou Bobo BARRY, Thierry FERRE, Djiby DIA, Aboubacar HEMA. 2022. Caractérisation des PME de transformation de l'anacarde au Sénégal et analyse des pratiques d'innovation. 50p.

MANE, Thierry FERRE, Djiby DIA, Aboubacar HEMA. 2022. Caractérisation des PME de transformation de la mangue au Sénégal et analyse des pratiques d'innovation. 53p.

Mor NGOM, Thierry FERRE, Djiby DIA, Aboubacar HEMA. 2022. Caractérisation des PME de transformation du riz au Sénégal et analyse des pratiques d'innovation. 55p.

Dr MEDAH Ignace, KIOGO Raymond, DOUGNON Gérard, HEMA Aboubacar. 2022. Rapport d'étude de la filière mangue au Burkina Faso.

Pape Bilal DIAKHATE et al. 2022. Diagnostic des Organisations Professionnelles et interprofessionnelles (OP et OIP) au Sénégal : cas des filières anacarde, mangue, arachide et riz.