JACQUES BOOYSEN

SENIOR DATA SCIENTIST · CITY OF JOHANNESBURG, SOUTH AFRICA · +27835085567

DETAILS

City of Johannesburg
South Africa
+27835085567
booysenjacques@gmail.com

LINKS

Linkedin

Portfolio

GitHub

LANGUAGES

Afrikaans

English

SKILLS

Critical thinking and problem solving

R Programming

Python Programming

Geospatial Data

Ability to Work Under Pressure

Self Starter

Rapid Prototyping

HOBBIES

Adventure Racing,
Ultra Trail Running,
Mountain Sports

PROFILE

I am a scientist at heart, passionate about Data Science, problem solving and software engineering. I am responsible for helping utilities and other companies interpret and manage data and solve complex problems using expertise in a variety of data niches often employing computer science, serverless cloud computing , machine learning, optimisation, modelling, statistics and advanced visualisations.

I have been involved in the development of Geo-based load sub classes , a market intelligence system with sentiment analysis using machine learning and integrated resource least-cost planning for energy trading market analysis on the Southern African Power Pool, including the impact of climate change models on this. I have produced day ahead hourly renewable forecasting models for IPP's. Solutions were developed using R, Python, specialised software and tableau.

My current role is technical lead in the disruptive innovation business unit at Enerweb, I provide Data Science as a service on the Python based in house SaaS technology stack Enerflow. We use AWS micro services architecture and a custom framework I developed that runs on top of Apache Airflow making use of AWS Lambda and allowing you to run Jupyter notebooks in the cloud from MS Excel. My strength lies in my ability to do rapid prototyping because of my strong software engineering background, and my passion for machine learning and constantly improving, doing things in a better way.

EMPLOYMENT HISTORY

Senior Data Scientist at Enerweb, South Africa
January 2019 — Present

Created renewable energy forecasting models and interactive tools for forecasting Solar PV, Wind and Hydro power plant outputs hourly using global circulation models and historic satellite data, including an interactive R shiny app applying the models bottom up for all IPP sites in South Africa to give a aggregated forecast for solar and wind plant generation. Used both Python and R and an XGBoost model.

Developed AWS micro services Python framework for our Data Science stack deployed using CI/CD and Infrastructure as code. Framework uses Amazon Apache Airflow (MWA DAGS), Lambda, S3, SQS, TexTract (OCR), Cloud-formation (AWS SAM) and the AWS miroservice framework. Developed data extraction and anomaly detection tool set for energy utility using k-means clustering in Python and statistics. Also working with ploomber and papermill to run Jupyter notebooks in production.

Developed hourly price forecasting models and visualisations for the day ahead SAPP power pool electricity market in Southern Africa, using random forest models in R and tableau.

Skills: Apache Airflow · Ploomber · Python · R · AWS · Jupyter · Tableau

Data Scientist at Enerweb
January 2012 — Present

Optimal regional geo based load forecasting (GLF), load subclass development finding sweet spot between error and model complexity, also a spatial buffering algorithm to calculate domestic building density estimates per spatial LSM. See my published papers on my portfolio site for more details.

Spatial interpolation of temperature using digital elevation model and comparing and evaluating various modelling techniques, a market intelligence system in R shiny for

network planning support and sentiment analysis using MongoDB and machine learning. Used full text and spatial search functionality.

Various projects using an electricity market model that was written in MESSAGE from the International Institute for Applied Systems Analysis (IIASA) to estimate least cost energy flows within and between the interconnected SAPP-EAPP power pools, given a range of scenario assumptions, A dynamic, bottom-up, multi-year energy system model based on the IRENA model applying linear & mixed-integer optimisation techniques was used. Results were shown over a medium to long term till 2040 and for different scenarios. Each country was modelled as a node interconnected by cross-border Tx Lines. This included an Inter-connector Impact Study of ANNA a 400kV inter-connector between Angola and Namibia, and another project for CRIDF using downscaled Climate Projections from NASA Earth Exchange (NEX) with the same energy model in a Monte Carlo Simulation. Both R Shiny and Tableau were used as the model interfaces. See portfolio above.

Various other research projects developed for Eskom research on our in house DIAS platform (before R Shiny existed), projects ranging from non-techinal losses, DSM impact studies and optimal power plant maintenace.

Skills: MongoDB · MESSAGE · Energy Modelling · Optimisation · R Shiny · Tableau · Python (Programming Language) · R (Programming Language) · SQL

Software Engineer at Enerweb
January 2002 — December 2009

Started as a Linux programmer and Back-end Specialist, doing various research projects in C and Perl.

I was also the solutions architect and main developer piloting the Virtual Power Station (VPS) scada solution for Eskom, this involved a distributed system running on Linux servers over corporate GPRS/4G network, with integration with Schneider Momentum PLC and Frequency Relays. The initial solution was written in Perl and C, with my own C implementation of Modbus to talk to the PLC and SPABUS to talk to the Frequency Relays. I also developed an IVR solution using JBoss and Linux asterisk. Also did some PLC programming. The initial solution has been re-written in C# by our dev team for productization.

Eskom decided to embark on a self developed, enterprise wide data warehouse, for financial, technical, HR and environmental data. This solution had to provide a "single version of truth", across all divisions, and had to be developed specifically according to Eskom business processes and technical standards. Immense technical complexity was overcome to produce a world leading technical solution. I was responsible for and developed a generic recursive roll-up engine as Java stored procedures that could navigate and extract multidimensional roll-ups over cubes and complex hierarchies in the data model.

Wrote a X.25 Protocol gateway using Linux C/Visual Basic and Open VPN, assisting Eskom Engineers to connect to substation equipment from the Eskom WAN.

I also wrote an authenticating firewall custom system for Eskom Telecoms, enabling control and secure access to the operational telecoms network via software that could only use RS232 ports.

Skills: Java · Perl · C (Programming Language) · J2EE Application Development · PHP · PL/SQL · Oracle Database · Modbus · SCADA · Programmable Logic Controller (PLC) · Linux

EDUCATION

B.Sc (Hons) Computer Science, University of Johannesburg
January 1998 — December 2001

Bachelor of Science (BSc) Applied Mathematics, University of Johannesburg
January 1998 — December 2001

PUBLICATIONS

Domestic building density estimates for network planning
April 2014

Domestic Use of Energy Conference (DUE) - see LinkedIn

Regional electricity load profile subclasses for network planning
August 2013

Industrial and Commercial Use of Energy Conference (ICUE) - see LinkedIn

CERTIFICATIONS

AWS Certified Solutions Architect – Associate
August 2022 — August 2025

R Programming
February 2015 — Present

Coursera Verified Certificates X9HT2EW4WW www.coursera.org

The Data Scientist's Toolbox
December 2014 — Present

Coursera Verified Certificates 2NZPCAYE53 www.coursera.org

COURSES

Water Evaluation And Planning System (WEAP) Intro, SEI — Stockholm Environment Institute
October 2020 — October 2020

REFERENCES

Marcus Dekenah from MD Consulting
marcus@mdekenah.co.za

Schalk Heunis from Vodacom
schalk.heunis@gmail.com

PROFESSIONAL MEMBERSHIPS

The South African Institute of Electrical Engineers , Johannesburg
April 2019 — Present

Member and presenter at the load research chapter, Presenter at SatRday 2018: Practical applications using R for spatial data visualisation,creation and manipulation