Julian Kaljuvee
kaljuvee@gmail.com – +44 7585 099 329 (m)

SUMMARY

- Full stack Data Scientist / ML / AI Engineer with over ten years of experience in statistics, data science, and software / data / ML engineering, primarily in financial services, semiconductor, media, retail / FMCG domains
- Strong knowledge of statistical methods, machine learning, artificial intelligence, and general software engineering techniques across a wide set of technologies and programming languages
- Professional work / consulting experience includes Goldman Sachs, Morgan Stanley, JPMorgan Chase, Capital, BoFA / Merrill Lynch, RBS, UBS, London Stock Exchange, Standard Chartered, HSBC, DBRS Morningstar, ARM Holdings, Nomad Foods, Nandos UK, IKEA, and Worldremit

DATA and SOFTWARE ENGINEERING

Languages:	Python, R, Java/Scala/Groovy, Solidity, Vyper
Cloud:	AWS, Azure, Google Cloud Platform (GCP)
Databases/Big Data:	PostgreSQL, Presto, Athenai, Redshift, BigQuery, MongoDB, Databricks, Snowflake, PySPark/SQLAlchemy, MS-SQL
Frameworks/ DevOps:	Kubernetes, Docker, Terraform
ML/AI libraries:	TensorFlow, PyTorch, SciKit, SparkML, spaCy, BERT, FB Prophet, FB Pytext
Visualisation:	Matplotlib, Seaborn, Plot.ly, Streamlit, Tableau, Apache Superset, Looker, Metabase

ML / AI / DATA SCIENCE and STATISTICS

ML / AI / DS:	Supervised learning including classification / regression and prediction / forecasting, Unsupervised and reinforcement learning, NLP, clustering, anomaly detection, ANN, GAN, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) / LSTM
NLP / NLG / NLU:	Stemming / lemmatising, Coreference resolution, Part-of-speech (POS) Tagging, Dependency Parsing, Named Entity Recognition (NER), Bag of Words (Tf-Idf), word embedding (Word2Vec), language models (BERT)
Data cleansing:	Pre-processing (Encoding, Training, Standardisation), Encoding, Missing Value Imputation, Model architecture design
Statistical techniques:	Regression (Linear, Kernel), Classification (Logistic Regression, Bayesian inference, Support Vector Machines, Decision Trees, Random Forest), Clustering (Nearest Neighbour), dimensionality reduction, principal component analysis, causal inference, Ensemble models
Time-series analysis:	univariate / multivariate analysis, seasonality adjustment, stationarity, co-integration, smoothing techniques, GARCH, ARMA / ARIMA
Generative AI:	LLMs (Cohere, OpenAI, Mistral, Gemini), RAGs, vector databases

PROFESSIONAL EXPERIENCE

Predictive Labs Ltd, (www.predictivelabs.co.uk) London, UK / Remote Mar 2020 - present
Data Scientist / ML / AI Engineer (Contract)

- **Vaxart (NASDAQ:VXRT)** – building bio AI and data science models for a US publicly traded biotech, including manufacturing optimization and using generative AI to drive insights
- **Denario / Pandadoc** (www.pandadoc.com) - pre series-A Fintech in AI / ML; B2B payments, embedded finance and revenue-based financing (RBF) space. Deployed LLMs RAGs for unstructured document processing
- **Worldremit** (www.worldremit.com) - revenue / sales forecasting, cloud migration, reconciliation of EOD and EOM finance activities for a global Fintech
- **IKEA** (www.ikea.com) - group wide KPIs and business strategy planning for IKEA / INGKA group
- **Nandos**, (www.nandos.co.uk) - sales / demand forecasting, data infrastructure development / data lake, POS, loyalty, rota, weather, wifi and geolocation and delivery data integration for a major restaurant chain
- **Nomad Foods**, (http://www.nomadfoods.com) - developed and productionised media / marketing mix model (MMM) based on the industry standard Nielsen model to measure advertising / media effectiveness spend (ad stock decay, competitor data), including Facebook Marketing API, Google Ad and AdWords API. largest frozen-food / FMCG retailer in Europe, marketing channels
- **ARM Holdings**, (http://www.arm.com) - development and productionisation of demand forecasting / royalty and non-royalty **revenue prediction model based on external (alternative)** data sets for a leading semiconductor design and manufacturing
- **Predictive Trade**, (https://predictive.trade) - development LLM-based news analysis and statistical arbitrage research programme (a side project)
- **Stack(s):** Git / Github, Airflow; Metabase; AWS, GCP, Azure; BigQuery, Redshift, Pytorch; MLOps: Vertex.AI, MLFlow, Wandb.ai

DBRS Morningstar, (http://www.morningstar.com) London, UK / Frankfurt, DE / Remote Jul 2019 – Mar 2020
Data Scientist (Contract)

- **Business Area:** structured finance and credit ratings
- **Responsibilities:** supporting development and deployment of EU RMBS rating models for a leading global financial analytics and ratings provider
- **Model building and deployment:** (i) data analysis / visualisation, (ii) data cleansing and transformation (data sets with over 100m records), (iii) model development, feature engineering and model evaluation with statsmodels, GAM and XGBoost, (iv) model deployment and productionisation with MLFlow (v) presentations to business stakeholders
- **Stack:** Git / Github, Domino Datalab, Jupyter, Jenkins; Python (Pandas, Scikit-learn, Statsmodels, Plotly, PySpark), MLFlow, Docker, Kubernetes, AWS (S3, Athena, Redshift), Databricks

Data Scientist / Engineer (Contractor)

- **UBS,** Data Scientist / Engineer, Market risk management (Jun 2018 – Jul 2019)
- **HSBC,** Data Scientist / Engineer, Market risk management (Jan 2018 –Jun 2018)
- **Standard Chartered Bank**, Data Scientist / Data Engineer, Market risk management(Mar 2016 – Dec 2017)
- **HSBC**, Date Engineer, Product Control (Nov 2015 - Feb 2016)
- **London Stock Exchange Group (LSEG) / LCH,** Data Scientist, Market risk (Jul 2014 – Nov 2015)
- **UBS**, Quantitative Analyst, IRD/FX/MM Front-Office (Jan 2010 – Jun 2014)
- **Royal Bank of Scotland**, Quantitative Analyst, Counterparty Credit Risk (Jun 2009 – Jan 2010)
- **ING Bank**, Quantitative Analyst, Counterparty Risk (Mar 2009 – Jun 2009)
- **Merrill Lynch/Bank of America**, Quantitative Analyst, Equity Exotic Derivatives (Mar 2008 – Feb 2009)
- **Barclays Capital**, Quantitative Analyst, Emerging Markets- Fixed Income and Commodities (Jan 2007 –Jan2008)
- **UBS Warburg,** Software Engineer, FIX Engine development (Jan 2006 – Dec 2006)
- **JPMorgan,** Quant Developer, Algorithmic trading (Jan 2004 – Dec 2005)
- **Goldman Sachs,** Quant Developer, Algorithmic trading (Jan 2005 – Dec 2005)

Harvard University, Cambridge, Massachusetts
Bachelors in Applied Mathematics
Course work included Economics, Applied Mathematics, Computer Science, Molecular Biology and Physics

Columbia University, New York, New York
Masters studies in Statistics / Operations Reasearch
Studies included Statistics, Probability Theory and Operational Research

OTHER / PERSONAL

LinkedIn:	https://www.linkedin.com/in/juliankaljuvee/
Github:	https://github.com/kaljuvee/datascience
Blog:	www.medium.com/@kaljuvee
Startups:	Co-founder of Bondora.com, EstateGuru.co, Investly.co and Landex.ai
Languages:	Estonian (native), English (fluent), Finnish (fluent), Russian (intermediate), French (intermediate), Swedish (beginner)

TECH SKILLS

Languages	
Python (Pandas, Numpy, SciPy)	5 Years
Java/ Groovy / Scala	5 years
R	5 Years
JavaScript	> 10 Years
ML / AI / Data Science	
Machine Learning / Deep Learning: TensorFlow, Theano, Keras, scikit-learn, SparkML, PyTorch, Prophet	3 Years
NLP: GloVE, spaCy, NLTK	2 Years
AutoML: SageMaker, Azure ML, Google AutoML	2 Years
Visualisation	
Matplotlib, Seaborn, Plot.ly, Tableau, Apache Superset, Power BI, Looker	3 Years
Devops / Data Scientisting / Dev Tools	
Airflow	3 Years
Terraform	3 Years
Jenkins	3 Years
Docker, Kubernetes	3 Years
Git / Github / Gitlab	5 Years
Jupyter Notebook, BeakerX	3 Years
Cloud Technologies	
Google Cloud Platform / GCP	3 Years
MS Azure	4 Years
AWS (EMR, Athena, S3)	4 Years
Middleware / Messaging	
Kafka, Kinesis, Spark / Databricks	4 Years
Elasticsearch	> 5 Years
Databases / ETL	
Redshift	2 years
Snowflake / DBT	2 years
Redis	5 Years
RDMBS: MySQL, PostreSQL, Presto, MS SQL Server	> 10 Years
BigQuery	2 Years
Presto	2 Years