

Patrick GREEN

Machine Learning Engineer | Data Engineer | Team Lead

 <http://patchgreen99.github.io>  [linkedin.com/in/paddy-green-a15b0aa6](https://www.linkedin.com/in/paddy-green-a15b0aa6)

 github.com/patchgreen99  bitbucket.org/patchgreen99 medium.com/patrick-green-78196

 +44 7542 648852  patchgreen99@gmail.com References available upon request

Passion for physics, computer simulations and computer vision with experience in feature engineering, stream processing, sports analytics and MLOps.

Skills

Education : BSc 1st in Computer Science and Physics from The University Of Edinburgh

Certifications : GCP Professional Data Engineer, GCP Professional ML Engineer, LookML Developer

Languages : Python, Scala, (C++, C, Java, JavaScript, Go, Typescript, Node, Flutter, Dart)

Frameworks : Spark, Spark Structured Streaming, Tensorflow, PyTorch, Akka Streams, Spring Boot, MLFlow

Cloud : AWS, GCP, Postgres, MySQL, Redis, BigQuery, Serverless, Kubernetes, Terraform, Firebase, AI Platform, AutoML, Confluent Kafka, EMR EKS, Apache Druid, Superset, Airflow, Beam, KSQL

Software + Hardware : NVIDIA, OpenVINO, CUDA, Arduino, FFmpeg, Looker, Jenkins, Github Actions

Experience

07/2022-12/2024 | Tech Lead, IMG Arena,

- As Tech Lead of the DataOps Team in IMG Arena I line manage 6 developers & data scientists and deliver innovations to enable fan engagement, personalisation and live stats for up to 4 million users (*The Open*). I also run our code review sessions, QA sessions, sprint planning, bug triage and daily stand ups
- **Productionised Real-time Models** Using Spark Structured Streaming, Scala, MLFlow and Python we host our own sport agnostic inference pipeline with which we have delivered models for shot quality and strokes gained. After we went live I have also refined our models to suit our clients for example we now derive additional course features from satellite images (Clere Golf), retrain and promote new versions for the PGA and DP World Tour. Click *here* for our shot view at The Open where our strokes gained model was live for 4 million users in 2023.
- **MLOps** Our team paved the way for data science in IMG Arena, using MLFlow and Airflow to create a pipeline that supports frequent retraining and optimization of our productionised GAM, XGBoost and Logistic models. In addition we released monitoring for data errors, concept drift, inference degradation and model performance for live models and testing on our historical time-series data.
- **Win Probabilities** We developed various ranking algorithms utilising our historical data and setup automated training and optimisations in airflow. Gathering probabilities from our real-time pipeline we setup a micro-service for simulating the likelihood of winning and other outcomes using statistical methods to adjust our predictions as new data arrived. We also acquired competitors in-play prices for accessing performance across the event.
- **AI Commentary** Our live commentary utilises an ensemble of models to weight individual events that maximise the user retention and click through rate seen on our *Mixpanel*. Using this heuristic our algorithm would choose the most relevant insight/comment to produce within seconds of that event happening. Visit our golf feed *here* and our MLS feed *here*.
- **Major League Soccer** Using a computer vision platform to generate tracking data at the stadiums we worked on data quality monitoring and alerting as well as models to derive possession and events from the spatial data so we could enrich it for analytics. *Here* are some insights we've generated.
- **Snooker Computer Vision** To collect more detailed data for snooker we developed and implemented a model for identifying ball location and colour from an on site camera. We used a mixture of mathematical techniques and ML for extracting the data from each frame and created a serverless pipeline for processing the frames and tools for calibration for the on-site technicians to use.

AWS | Kubernetes | Terraform | Python | Scala | Akka Streams | Spark Structured Streaming | Druid | Superset | Airflow | Github Actions | MLFlow | EMR on EKS | Redis | Confluent Kafka | KSQL | Bigquery | PyTorch | OpenCV | FFMPEG | statsmodels | sklearn | pygam | R | pymcmcstat | pymc

10/2021-07/2022	<p>Senior ML + Data Engineer, IMG Arena,</p> <ul style="list-style-type: none"> ➤ Data Lakehouse To enable data products, analytics and ML I designed and implemented a fault tolerant data lakehouse that can model golf, basketball and football. We manage our own EKS deployments of Apache Druid, flask and redis for hosting and serving our data products. We also manage our own helm spark streaming jobs and ksql clusters to identify labelling errors, apply corrections, union with historical data and enrich data with other sources so we can produce a single live query able table that's fault tolerant and can scale. In 2023 we released new insights from our Lakehouse including a broadcasting tool for the <i>DP World tour</i>, a <i>course difficulty tool</i>, a <i>tournament winners dashboard</i> and <i>league leaders widgets</i> for Euroleagues. <p> AWS Kubernetes Terraform Python Scala Akka Streams Spark Structured Streaming Apache Druid Apache Superset Apache Airflow Github Actions Jenkins MLFlow EMR on EKS Bigquery </p>
03/2021-10/2021	<p>Video Streams Engineer, IMG Arena, https://www.imgarena.com/sports-betting/live-streaming/</p> <ul style="list-style-type: none"> ➤ IMG Arena delivers video streams for over 300 sports events per week ➤ Live Video Streaming Platform I worked on the CDN in Akamai setting up the mapping between our publishing points and hostname for our HLS streams. I also setup the delivery reporting platform which processed 1TB of user sessions a day for a customer analytics query tool that we manage for every client. Before going live I also worked with the team to ensure customers would have access to the legacy platforms historical data going back five years both in Apache Druid and MySQL. <p> AWS GCP Kubernetes Scala Java Akka Akamai Bigquery Apache Druid MySQL FFMPEG </p>
10/2018-03/2021	<p>Consultant, Servian, https://www.servian.com/</p> <ul style="list-style-type: none"> ➤ As a Google Certified Data Engineer, a Machine Learning Engineer and a LookML Certified Developer my engagements have focused on Data Analytics, Data Processing, Data Science and general Google Cloud work. I also ran a live demo of Apache Beam at a Google event that you can watch here. ➤ Customer retention in Westfield with Computer Vision, Using shopping mall and store surveillance cameras we developed a product that empowers a store owner. Using event based tracking, we log impression data on items in the store. We also allowed the store to evaluate its footfall, traffic and interactions and how well each shoppers need was met by the staff. Finally we experimented with a completely contact less checkout system similar to that of Amazon Go's. ➤ Sports Analytics, I used Bigquery to process the GPS and video data and then built a web app using Looker for analysing Rugby League Games. The web app also served the results of a BigQueryML model to deliver an automatic labeling solution that could categorize events based on the GPS data from each player in real time. I also delivered a webinar to promote our partnership with Looker here. <p> Python OpenCV OpenVINO NVIDIA Tensorflow Google Cloud Apache Beam BigQuery AutoML Looker </p>
05/2016-08/2018	<p>Backend Developer, TVSquared, https://tvsquared.com</p> <ul style="list-style-type: none"> ➤ TVSquared hosts online portals to help brands optimise their TV advertising spend. ➤ Modeling an Ads retention, Working with Spark I implemented an algorithm that could identify the household of any visiting user received from our Javascript web tracker. I utilized Postgres to store the user and household identities. I also implemented several asynchronous loaders that run on TVSquared's multi threaded task queue and apply further tracking and demographic data from a Cookie Syncing Partner to our modeling. https://tvsquared.com/tvsquared-liveramp-partner <p> Python Spark PySpark Mongo PostgreSQL AWS TeamCity </p>
Personal Project	<p>Ranking Surf Conditions using webcams, Computer Vision, https://patchgreen99.github.io</p> <ul style="list-style-type: none"> ➤ The application assisted in the ranking of surfing conditions across beaches by analysing wave and surfer behavior using live, low resolution web cameras and delivering the results to an app. I trained a model using preprocessed web camera recordings to infer the quality of the surf. I also used transfer learning to build models for surfer detection and used various computer vision algorithms to stabilize the cameras and geolocate the surfers. <p> Python OpenCV TensorFlow CUDA FFmpeg OpenVINO Google Cloud Cloud Tasks Cloud Functions Firebase Cloud Run AI Platform Dart Flutter </p>