Laurie Lugrin
R&D engineer

laurie.lugrin@gmail.com
https://github.com/Lugrin
07760 588 345
London

Skills and Interests

Programming	Python (NumPy/pandas, sklearn, nltk, gensim), Scala/Spark, Bash, C++
Software engineering	Algorithms & data structures, git, Linux
Machine learning	Forecasting, Recommender systems, probabilistic programming
NLP	Named entity recognition, entity linking, word embeddings

Experience

Data Scientist at The Very Group, London, UK   Nov 2017 – today (3.5 years)

The Very Group is the second largest online retailer in the UK. I contributed to many projects, from customer intelligence to logistic optimisation.

- Unconstrained demand estimation using a Gradient Boosted Trees model to account for product unavailability. This was used as preprocessing before we run our demand forecasts.
- Customer feedback analysis using topic modelling and phrase modelling. The results were made of time series, word clouds and representative customer comments. This tool provided insight quickly on all 8 kinds of customer surveys so that we can take actions before the text classification is trained and deployed.
- Search auto-complete suggestions based on frequent searches, using equivalence classes to remove pluralisation and word-split duplicates, ensuring that we show relevant and diverse suggestions to the customer. Words that do not appear in any product description were flagged as potentially inappropriate and eliminated.
- Search-term classifier using an LSTM neural network with word embeddings. The output provided insight on the customer demand to the trading team.
- Delivery cost model that includes the risk of loss/damage and customer dissatisfaction so that we can choose the most cost-effective carrier service for each delivery.
- Real-time alerting system for abnormal demand so that we can get enough stock in time.
- Advisory role in various projects, including to the chatbot team, implementing text classification and sentiment analysis of customer surveys, and using AWS services to help the data science team develop and productionise their work.
- Organising and speaking at internal knowledge-share sessions.

R&D NLP engineer at Idioplatform, London, UK   Jan 2015 – Oct 2017 (3 years)

Idioplatform helps brands better understand their prospects and their online content, in order to build a marketing strategy driven by data. As part of the research team, I worked on the semantic text-analysis engine at the core of their content-management system.

- State-of-the-art named-entity recognition and disambiguation system, beating dbpedia-spotlight, Zemanta and Alchemy in F1-score on academic datasets.
- Evaluation: precision/recall evaluator with flexible string matching; non-regression tool with detailed feedback for identifying patterns in errors; automated test set creation for a specific

domain (fashion, finance); grid-search-inspired hyper-parameter optimisation tool with visual output to provide insights on the role of each parameter.

- Ontology incorporating different data sources, regularly updated with new topics from open-source knowledge bases, so that we can report on the latest topics, for example new technologies and current affairs.
- Sanity-check tool for our ontology, identifying orphan or duplicated entities as well as anomalies (rule-based), so that we are confident when releasing updates.
- Data-pipeline automation. We can rebuild any intermediate or production dataset in one command.

Risk software engineer at RenaissanceRe, Dublin, Ireland   Jan 2012 – Dec 2013 (2 years)

RenaissanceRe is a reinsurance company with a large volume of contracts signed daily. As part of the backend team, I made sure the analysts had the best tools to understand and quote their deals.

- Rewrote the insurance risk-estimation software using Monte-Carlo methods, improving speed, maintainability and extensibility.
- Designed a data format for contract terms that is intuitive to analysts and has a straightforward implementation. Proved the equivalence of the two representations. Wrote a migration tool to transform the legacy files.
- Developed a new software that generates different scenarios of human errors and estimates their impact, accounting for various parameters such as profession, region and type of insurance; from requirement analysis to tests and integration.

Software engineer at Moody's analytics, Montbonnot, France   June – Aug 2011

Developed a rule-based system that determines the safety-net threshold for bank loans according to regulations.

Research assistant at Verimag Lab, Grenoble, France   Oct 2009 – Dec 2010

Formulated a method for comparing energy consumption models of wireless sensor networks. Performed a case-study using the datasheet of the embedded radio device CC1100.

Computer science tutor at Joseph Fourier University, Grenoble, France   2008 – 2010

Led 150 hours of tutorials and practical labs: C, algorithms, formal languages, automata theory.

Research intern at the University of Toronto, Canada   May – Sep 2008

Education

Continuous Learning

- Personal projects: IRC bot that makes rhymes and funny remarks, sunrise video generation, bot that tries and survives in a multi-agent iterated prisoner's dilemma environment.
- Conferences: Speaker at PyData Paris 2016 conference.

M.Sc. on Computer Science, Joseph Fourier University, Grenoble, France   June 2009

Minor in Artificial Intelligence, with high honours (80$^+$%)

Hobbies

Hiking, kickboxing, table tennis, badminton, ski, via ferrata, juggling, cycling, guitar, crochet.