Laurie Lugrin

Senior Data Scientist / ML Software Engineer

laurie.lugrin@gmail.com
https://github.com/Lugrin
Remote or London onsite

Interests

Software engineering	Python (pandas, sklearn, nltk), Algorithms & data structures, git, Linux
Machine learning	NLP, Time-series analysis, Probabilistic Programming

Experience

Senior ML Software Engineer at SenseOn, London, UK Jan 2022 – present (1 year 8m)

SenseOn is a cyber-threat detection platform. As part of the ML team, I worked with security analysts to build several detectors of suspicious activity.

- Detecting obfuscated Powershell scripts. I combined a regex approach to cover simple obfuscations, and trained an LSTM on a publicly-available labelled dataset to cover more advanced obfuscation techniques, resulting in an F2-score of 87%.
- I designed a potentially malicious activity detector which combines a convolution-based algorithm that predicts users' working hours and a classifier of interactive programs.
- Detecting malware beaconing: I implemented a tool to detect suspicious programs that attempt network connections at semi-regular intervals.

Senior Data Scientist at The Very Group, London, UK Nov 2017 – Oct 2021 (4 years)

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

- Unconstrained demand estimation using Gradient Boosted Trees to account for product unavailability.
- Stock intake prioritisation using Convex Optimisation to maximise availability and revenue while meeting warehouse capacity constraints.
- Customer-feedback survey analysis and reporting using Topic Modelling and Phrase Modelling.
- Search auto-complete suggestion engine based on frequent searches, removing duplicates such as plurals, word-split variants and misspellings, ensuring that we show relevant and diverse suggestions to the customer.
- Search-term classifier using an LSTM neural network with word embeddings. The tool provides insights on customer demand trends to the trading team.
- Delivery cost model including risk of loss/damage and customer satisfaction which decides the most cost-effective carrier service for each delivery.
- Advisory role on various other projects, including chatbot evaluation, text classification and sentiment analysis of customer surveys.
- Mentoring colleagues.
- 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 benchmark 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 hyper-parameter optimisation tool with visual output to provide insights on the role of each parameter.
- Knowledge graph in neo4j, incorporating different data sources, regularly updated with new topics from open-source knowledge bases, to pick up the latest topics, for example new technologies and current affairs.
- Sanity-check tool for our knowledge graph, identifying disconnected or duplicated entities as well as anomalies using various graph algorithms such as cycle detection.
- Data-pipeline automation using Luigi and AWS. 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 developed tools to help analysts 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.

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.

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

Formulated a method for comparing energy consumption models of wireless sensor networks.

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

Personal projects

- AI that tries and survives in a multi-agent iterated prisoner's dilemma environment.
- IRC AI that makes rhymes and funny remarks.
- Speaker at PyData Paris 2016 conference.

Education

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

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

Hobbies

Hiking, kickboxing, badminton, skiing, via ferrata, crochet, pottery.