Yunfei (David) Bai, MASc

Cell: (905) 962-2557 | [baiyfei@hotmail.com] | [LinkedIn Profile: Yunfei Bai] | [GitHub: Yunfei Bai]

Principal Data & ML Engineer | Hands-on Engineering Leader

Summary of Qualifications

- Accomplished Principal Engineer and hands-on technical leader with 10+ years of experience in architecting, building, and scaling data-intensive machine learning applications and platforms from concept to production.
- Proven expertise in end-to-end ML lifecycle and MLOps practices: leading the design of scalable data pipelines, feature engineering, model development (NLP, Deep Learning), deployment (Docker, Kubernetes), monitoring, and CI/CD automation.
- Deep technical proficiency in Big Data technologies (Spark, Databricks, Elasticsearch ELK stack), cloud platforms (AWS, RedHat OpenShift), and programming languages (Python, Scala, SQL) for high-volume data processing and analytics.
- Strong background in data engineering and architecture: designing enterprise data warehouses, developing complex ETL/ELT processes, and building robust data governance and QA frameworks.
- Demonstrated ability to set technical direction, mentor high-performing engineering teams, and collaborate cross-functionally with data scientists, business leaders, and IT to translate complex business needs into innovative technical solutions.
- Holder of advanced certifications in Artificial Intelligence and Data Science from the University of Toronto, complementing a Master of Applied Science in computational modeling.

Professional Experience

Royal Bank of Canada (RBC) | Toronto, ON
Senior Manager, TI IT Risk Analytics | Feb. 2023 – Present, **Data Scientist** | Mar 2020 – Jan 2023

- Technical Leadership & Strategy: Led the end-to-end architecture, design, and development of a next-generation TI Risk Analytics platform from scratch. Defined the technical vision, selected technologies (Elasticsearch, Kubernetes, Python), and established best practices for data engineering and MLOps.
- Hands-on Data & ML Engineering: Built and automated scalable data pipelines ingesting from 15+ sources (handling millions of records/daily) for KPI/KRI insights. Architected a centralized data repository and implemented automated data validation and QA processes, improving data reliability and reducing manual effort by 90%.
- End-to-End ML Product Development: Conceived, developed, and productionized multiple NLP and machine learning models to address key business risks:
    - Change Risk Mitigator (Sage Avert) with AI Ops team: Led the development of an ML model integrated with ServiceNow that predicts high-risk changes, reducing change-induced P1/P2 incidents from 8% to 2% and saving millions annually.
    - Role-Based Access Control (RBAC) Model: Architected an unsupervised learning algorithm to cluster 88K+ entitlements into roles, automating access attestations and resulting in cost savings of ~$100K per implementation phase.
    - RCSA Classifier Model: Built and deployed an NLP classification model on OpenShift to automate the risk categorization of 1000+ items, reducing a manual 2 expert task to a 15-minute automated process.
- Cross-Functional Partnership & Mentorship: Partnered with senior executives, risk teams, and IT to align projects with strategic goals. Mentored and coached a team of data engineers and analysts on software engineering best practices, code reviews, and advanced data techniques.

Tools: Python, Keras, TensorFlow, Spark, Docker, Kubernetes, RedHat OpenShift, Elasticsearch ELK Stack, Git, Jira

Bell Canada | Toronto, ON
Data Science Manager | Nov 2019 – Mar 2020, **Data Science Developer** | Jan 2017 – Nov 2019
- o Managed the portfolio and resources for a team of 5 senior developers on BI and ML projects, including Campaign Intelligence and Customer Personalization.
- o Designed and implemented an automated Machine Learning Factory framework to operationalize predictive models. This included creating data mart architectures, building ETL pipelines with SAS, and developing performance monitoring dashboards in Tableau.
- o Enabled the automated production scoring of models, significantly improving the speed from data-to-insight for customer experience analytics.

Technical Skills

- Programming & Data: Python (Expert), Scala, Spark (SQL, MLlib), SQL (T-SQL, Teradata), SAS, Bash
- Machine Learning: TensorFlow, Keras, Scikit-learn, NLP, Deep Learning, Reinforcement Learning, MLOps, Model Deployment & Monitoring
- Big Data & Cloud Tech: Spark, Databricks, Elasticsearch ELK Stack, Kafka, Airflow, Docker, Kubernetes, RedHat OpenShift, AWS (S3, EC2, ECS)
- Data Engineering: Data Warehousing, ETL/ELT, Data Modeling (Star/Snowflake Schema), Data Governance, Data Pipelines
- Tools & Platforms: Git, Jira, Tableau, Power BI

Certification

Certificate of Artificial Intelligence | University of Toronto | 2020
· Intelligent Agents & Reinforcement Learning (95%), Deep Learning (94%)
Certificate of Data Science | University of Toronto | 2019
· Machine Learning (98%), Big Data Management (98%), Statistics (99%), Foundations of Data Science (94%)

Education

Master of Applied Science (MASc) | McMaster University | Hamilton, ON | 2012