Jessica Claire

100 Montgomery St. 10th Floor
(555) 432-1000 - resumesample@example.com

Website, Portfolio, Profiles

- https://github.com/Claire007
- http://plotme.org/
- http://snowflect.com/

Education

11/2011 **Ph.D.**: **Materials Science & Engineering**
McGill University - Montreal, QC

08/2007 **Master of Science**: **Materials Science & Engineering**
McGill University - Montreal QC

Certifications

Project Management Professional (PMP), PMI
Six Sigma Black Belt (CSSBB), ASQ
Professional Engineer (P.Eng.), PEO

Skills

- ML models (Regression, Classification, Clustering)
- Feature engineering (Data Cleaning, PCA, VIF, Lasso, SMOTE)
- ML deployment (Hyperparameter tuning, Cross validation, ML Pipelines, Streaming Analytics)
- ML lifecycle (MLflow tracking, registration)
- ML framework (R, Scikit-learn, PySpark)
- TS forecasting (Prophet, ARIMA)
- DL (TensorFlow)
- NLP (vectorizer, word embedding - transfer learning, sentiment analysis - VADER)
- Data visualization (Tableau, Power BI, Qlik Sense, OAC, Microstrategy, R-shiny, Plotly)
- Data Engineering (Databricks, Structured Streaming, Hadoop ecosystem, Teradata, DB2/Oracle, EDW, Cloud migration)
- AWS (EC2, S3, Redshift, SageMaker, Lambda, EventBridge)
- Azure (Data Factory, SQL, ML Studio)
- GCP (GCS, BigQuery, Cloud Composer, Cloud Function)
- DevOps platform (GitLab)

Work History

08/2019 to Current **Consultant**
Deloitte – Florham Park, NJ

- **Client: CIO** (Oct 2022 - Present)
- Modernizing Endpoint Security & Operations Infrastructure Project
- ***Client: Healthcare provider*** *(Sep 2021 - Aug 2022)*
- Provided development and leadership support to end-to-end execution and validation of data migration from legacy on-prem (oracle, DB2) to cloud using GCP services (Cloud Function, Cloud Composer, GCS etc.), Databricks. Databricks structured streaming is used to move CDC data (Kafka). It , also, involves configure, sizing and optimization of Databricks Clusters.
- Develop machine learning capabilities on streaming data (streaming analytics) using PySpark/MLlib and ML Pipelines. This is to provide Realtime prediction capability (i.e. member outreach, flight risk etc.) on large volume of data.
- ***Client: Healthcare provider*** *(Feb 2021 - Sep 2021)*
- Develop NLP based machine learning and deep learning models (TensorFlow) to analyze and predict claims outcome (Azure/Databricks). Automated pipeline using HIVE metastore and Databricks Workflow.
- Provide data and analytics support in Risk Adjustment and CMS Star Quality (experiences of Medicare beneficiaries). Work involves understanding health insurance risk and quality concepts, Data transformation (Azure/Databricks), automated pipeline (Databricks Jobs) and load (ArangoDB) for frontend D3.js reporting.
- ***Client: Global Pharmaceutical*** *(Sep 2019 - May 2020)*
- Capacity modeling of clinical development network using process prediction (Bayesian regression) and discrete event simulation (DES). Models include development network of oral solid dosage (OSD) drug product, small molecule and biologics drug substance.
- Training data prepared and automated using AWS Redshift, SageMaker, Lambda and EventBridge. Frontend developed and deployed using R-Shiny and Domino Data Lab.
- **Client: Postal Services** (Jul 2020 - Sep 2020)
- Develop end-to-end analytics and modeling strategies based on current processes and client's business need. This involves connecting data sources to visualization platforms using database connectivity or REST API.
- Develop multi-class text classification model using Scikit-Learn on survey data (SurveyMonkey).
- **Client: Natural Resources** (Sep 2020 - Feb 2021)
- Implemented automated Data Pipeline using SAP BW, Power BI, powerapps and Data Gateway. It provides real-time reporting of financial transactions, cost models and cost forecasting to the stakeholders.
- Utilize Azure ML Studio to develop model (multiple linear regression (MLR)) to predict cost per operating unit and integrate with Power BI.

04/2014 to 08/2019 **Continuous Improvement Lead**
Tenneco Automotive – Elizabeth, New Jersey

- Develop end-to-end data solution using R-based data acquisition (on-premise SQL server, OSIsoft PI, LIMS), transformation, predictive modeling (product classification, production outputs), visualization (R-shiny) and deployment (LAN server).
- Implemented Power BI analytics and reporting capability.
- Develop engine failure model using Mine Haul Truck data for Hortonworks Hadoop environment (PySpark/ MLlib).
- Time series forecasting of sales/demand and DES model development to determine EOQ.

11/2011 to 04/2014 **Process Engineer**
Glencore Plc – Sudbury, Ontario

- Statistical process control (SPC) and process capability analysis using Matlab.
- Copper-Nickel separation process optimization using predictive modeling (multiple linear regression) and design of experiments (DOE).
- Implement maintenance contract management best practices by improved scheduling using MS Project, tracking progress and conducting lessons learned.
- Develop maintenance KPIs using SAP and VBA to track PMs and cost overruns. Facilitate plant wide idea generation.

Public Applications

- Web app to develop machine learning models (http://plotme.org/mlearn/)
- Processing Social Media (Twitter, Reddit etc.) feeds using Databricks, SQL and Data Visualization pipeline (https://snowflectanalytics.shinyapps.io/socialNet/)