Eric Keniuk

✉ eric.keniuk@gmail.com 📱 (224) 639-3648 in https://www.linkedin.com/in/erickeniuk/ ⬡ https://erickeniuk.github.io/

SUMMARY

Master's in applied economics and bachelor's in mathematics. Expertise in financial, telecom & sports analytics big data with experience in USG. Highly proficient in both qualitative and quantitative analysis. Skillset: Hadoop, Linux admin, Spark Scala, PySpark, Python, Java, Javascript, C, SQL, Unix / Bash, shell scripting, Linux, Networks, Firewalls, Git, cloud & distributed computing, ETL, Airflow Docker, Agile principles (Scrum), Machine Learning, Unsupervised Learning

EDUCATION

Master of Arts in Applied Economics

University of Alabama • Tuscaloosa, AL • December 2019 • GPA: 4.0

Bachelor of Arts in Mathematics and Economics

University of Alabama • Tuscaloosa, AL • May 2019 • GPA: 3.86

EXPERIENCE

Chicago Cubs

Data Engineer, Baseball Systems - Research & Design   **March 2024 - Present, Remote**

• New role: baseball systems, player data, stats, high speed footage, predictive analytics, CI/CD & MLops (AWS stack)

T-Mobile

Data Engineer   **March 2023 - March 2024, Bellevue, WA**

• ETL pipeline builds within MS Azure standing up resource groups (ADF, data sinks, key vaults, storage accounts,etc)

• Development in Agile methodology within Python data migration framework to allow easy user configuration and deployment; Support for deployment and bug troubleshooting; GPG blob streaming encryption automation

• Databricks notebook development and workflow automation scheduling for readStream/writeStream data migration of Delta Tables to Snowflake including automated Voltage encryption translation

Early Warning/Zelle

Sr. Data Scientist - Data Platform   **May 2022 - March 2023, Seattle, WA (Remote)**

• Developer (prototype DEV/QA) for synthetic identity fraud product in Spark Scala, version control (Git) w/ Artifactory deployment; led communications; regression, integration testing, automated ETL pipeline (Python, Unix)

• Developer on consumer behavior Spring API, process Solr, Hbase data and regression test; Unix pipeline automation. Contributed to retroactively score API requests; PO for full-stack internal Spring app for Data team

• Lead for transition of 5 teams to new on-prem data center/cluster: Built virtual environment (Windows, Linux), established firewall openings, installed applications/software, configured Hadoop 3.x, resolved/refactored script errors

Data Scientist   **September 2020 - May 2022, Chicago, IL**

• Automation of daily processes– data ETLs, report aggregation, and model performance and data drift monitoring via Bash, Python, Spark Scala, R; RSA keys and SSH, employ DataRobot, and Universal Controller system automation

• Feature engineered +800 features via Spark Scala, Git for new payments ML model

• Model validation, performance & governance reporting via Hive, SAS, Python. Automated processing pipelines, cut lead times by more than 50%. Enhanced SpringBoot PROD API to run concurrent instances and run more reliably.

PROJECTS

Synthetic Identity Fraud Attribute Product – Lead Development Internal Testing Code

• Attribute product to create input data for external modeling & rules frameworks for identifying fake identities

Payments Machine Learning Fraud Model w/ Minimal Inputs – Feature Engineering

• Feature eng. in Spark Scala of payments fraud ML model for targets with no contributed attributes data

Real-Estate Investor Effects on Primary School Outcomes – Master's Thesis

• R language to extract & transform 3MM observations of home sales from 1952 to 2015, identify small, medium, and large investors by home ownership volume and model impacts on educational outcomes: https://erickeniuk.github.io/