

Juraj Macko

Dátový inžinier PySpark

✉ juraj.macko@example.sk ☎ +421 900 118548 📍 Prievidza

👤 Zhrnutie

PySpark dátový inžinier so skúsenosťami s priemyselnými časovými radmi. Pripravuje reprodukovateľné dátové sady pre údržbu strojov a kontroluje, či modely pracujú s úplnými vstupmi.

🏢 Pracovné skúsenosti

- **Dátový inžinier PySpark**
Banícka Energia 1/2022 - Súčasnosť
 - Prepracoval PySpark dávky pre 240 miliónov vibračných meraní denne; s 3 ML inžiniermi skrátil aktualizáciu modelu zo 74 na 19 minút.
 - Zmenil join kľúče a cache pre 28 Spark úloh; náklady na výpočet klesli o 31 % bez zmeny počtu spracovaných strojov.
 - Pridal kontroly chýbajúcich vzoriek z 11 liniek; plán údržby prestal používať neúplné okná v 96 % behov.
 - Zapísal 5 TB surových signálov do Delta Lake s časovými oddielmi; vyšetrovanie poruchy sa skrátilo z 90 na 17 minút.
- **Dátový inžinier**
Hornonitrianske merania 4/2018 - 12/2021
 - Vytvoril Airflow pipeline pre 16 senzorových brán; 4 technici dostali ranný prehľad pred siedmou v 98 % pracovných dní.
 - Odstránil duplicity podľa času a sériového čísla v 72 miliónoch udalostí; falošné servisné tikety klesli o 39 % za štvrtrok.
 - Nastavil MLflow verzie datasetov s tromi výskumníkmi; 18 tréningov bolo možné zopakovať s rovnakým vstupom.
 - Zjednotil jednotky vibrácií z 8 typov snímačov; nesúlad reportov medzi zmenami klesol z 11 % na 2,4 %.

<> Projekty

Predikcia servisných zásahov
Dátový produkt pre oblasť vibrácie zariadení a prediktívna údržba, so spravovanými zdrojmi, verzovanými transformáciami a kontrolovateľnými pravidlami kvality.

PySpark
Delta Lake
MLflow
Airflow

🎓 Vzdelanie

Ing. v odbore automatizácia a riadenie
Univerzita Alexandra Dubčeka v Trenčíne
9/2014 - 6/2018

★ Zručnosti

Platformy a programovacie jazyky

PySpark, Delta Lake, MLflow, Airflow, Azure
Blob Storage

Kvalita a prevádzka

Modelovanie dát, Testovanie, CI/CD,
Monitorovanie, Dokumentácia

Spolupráca

Zber požiadaviek, Revízia kódu, Riešenie
incidentov, Práca naprieč tímami

🇸🇰 Jazyky Slovenčina Materinský jazyk
Čeština C1

**Databricks Certified
Associate Developer for
Apache Spark**
Vydávajúca organizácia
2025

🏆 Certifikáty

**Workshop kvality dát:
vibrácie zariadení a
prediktívna údržba**
Slovenská komunita dátových
inžinierov
2023