

Amalie Foss

PySpark Data Engineer

✉ amalie.foss@example.no ☎ +47 94307852 📍 Gjøvik

Sammendrag

PySpark Data Engineer med erfaring fra sensordata og maskinlæring. Forbereder stabile features med dokumentert lineage slik at modeller kan evalueres på samme grunnlag i utvikling og drift.

Arbeidserfaring

PySpark Data Engineer Innland Sensorikk · Gjøvik	Jan 2022 - Nåværende
- Transformerte 640 millioner vibrasjonsmålinger i Structured Streaming for 12 produksjonslinjer; modellteamet mottok features 14 minutter raskere etter at state ble redusert. - Optimaliserte skewed joins med salting og AQE på 9 TB Delta-data; full retrening gikk fra 6 timer til 2 timer og 18 minutter. - Publiserte versjonerte features gjennom MLflow og Delta Lake for fem modellutviklere; forskjell mellom trening og produksjon falt fra 8,4 til 1,2 prosent. - La til watermark og senhendelsesrapport for Kafka-strømmen sammen med fabrikkdrift; forkastede målinger gikk ned 72 prosent.	
Data Platform Engineer Innland Sensorikk Digital · Gjøvik	Apr 2018 - Dec 2021
- Bygget PySpark-jobber for batchdata fra 85 CNC-maskiner; prosessingeniørene fikk ukentlig verktøyslitasjeanalyse med 93 prosent komplett dekning. - Opprettet testdatasett med 14 kjente sensorfeil og integrerte dem i CI; regresjoner ble fanget i 17 av 19 pull requests. - Flyttet 3,2 TB råfiler til komprimert Parquet og beholdt kontrollsummer; lagringsbehovet ble redusert med 41 prosent. - Samarbeidet med data scientists om definisjon av 32 aggregater og skrev lineage-notater; modellvalidering tok 1 dag i stedet for 4.	

Prosjekter

Feature store for maskinsensorer	Et produksjonsnært dataprodukt for industrielle sensorer og feature engineering, med sporbare transformasjoner, tester og driftsrutiner.	PySpark	Delta Lake
		Structured Streaming	Kafka

Ferdigheter

Plattform og språk

- PySpark
- Delta Lake
- Structured Streaming
- Kafka
- MLflow

Datakvalitet og drift

- Datamodellering
- Testing
- CI/CD
- Overvåking
- Dokumentasjon

Samarbeid

- Kravavklaring
- Kodegjennomgang
- Hendelseshåndtering
- Tverrfaglig levering

Utdanning

Høgskolen i Østfold	Aug 2014 - Jun 2018
Master i industriell matematikk at industrielle sensorer og feature engineering	

Sertifiseringer

Databricks Certified Data Engineer Associate

Sertifiseringsorganisasjon

Mar 2025

Praktisk kurs i datakvalitet

Norsk Dataforum

Oct 2023

Språk

Norsk • Morsmål

engelsk • Profesjonelt nivå