

## CONTACT

🏠 Helsingborg, Sweden  
☎ 073 71 68 622  
✉ anas-mofleh@hotmail.com

🌐 <https://anasmofleh.github.io/>  
in <https://www.linkedin.com/in/anas-mofleh>  
🐙 <https://github.com/AnasMofleh>



## SKILLS

Data engineering:  
Datamodellering, data  
warehouse, pipelines,  
datakvalitet, business  
intelligence.

AI/ML:  
Bildbehandling, NLP,  
transformer-modeller,  
molnteknologier.

**Software developing:**  
Projektledning, agil utveckling,  
multitrådad programmering.

## TEKNISKA KOMPTENSER

Mycket goda kunskaper i:  
Python, Java, Scala, Git, MySQL,  
PostgreSQL, PySpark, Hadoop,  
Terraform, Azure, Microsoft  
Fabric, Azure Data Factory,  
Databricks, Snowflake, DBT,  
Airflow, Luigi, AWS S3, Gradle,  
Linux, REST API, MATLAB, Jira,  
Confluence, Excel,  
nätverksprotokoll (IP), Tableau,  
Power BI.

Goda kunskaper i:  
Docker, Kubernetes, GRPC,  
Haskell, R, MiniZinc, C#, C,  
C++, HTML

# ANAS MOFLEH

## DATAENGINEER MED AI/ML BAKGRUND

En nyckelperson i en BI-enhet, med rollen som Data Engineer. Ansvarig för att skapa och övervaka datarör (pipelines), konstruera dataflöden med hög datakvalitet och tillhandahålla BI-insikter. Stöttar beslutsfattande genom att omvandla data till affärsvärde. Har även kunskap att tillämpa AI för att ytterligare förädla datainsikter.

## ARBETSLIVSERFARENHET

### DATA ENGINEER

**Region Skåne | 2023-06 - Pågående, Lund**

Ett EU finansierat projekt som heter "Datasjö för hälsodata"

### DATA ENGINEER / BI DEVELOPER

**Sinch AB | 2021-10 - 2023-06, Malmö**

Skapade och underhöll pipelines samt utvecklade och förvaltrade dashboards inom en BI-enhet, vilket stödde sälj- och ekonomiavdelningar i deras beslutsfattande.

### MACHINE LEARNING ENGINEER

**Sinch AB | 2021-05 - 2021-10, Malmö**

Implementerade och underhöll avancerade transformerbaserade modeller för sentimentanalys.

## UTBILDNING

### CIVILINGENJÖR I DATATEKNIK – AI OCH MASKININLÄRNING

**Lunds Tekniska Högskola | 2017 - 2021, lund**

M. SC. i Artificiell intelligens and Maskininlärning

## CERTIFIERINGAR

- **2025-02** **FABRIC DATA ENGINEER ASSOCIATE**  
Microsoft
- **2024-10** **GITHUB ACTIONS**  
GitHub
- **2024-08** **DATABRICKS CERTIFIED DATA ENGINEER ASSOCIATE**  
Databricks
- **2024-07** **GITHUB FOUNDATIONS**  
GitHub
- **2024-06** **DATABRICKS LAKEHOUSE FUNDAMENTALS**  
Databricks

Grundläggande kunskaper: Snowpipe, Arduino, JavaScript, React Native, AWS Lambda, Kafka.

SPRÅK

**Arabiska:** Modersmål.  
**Engelska:** Professionell nivå  
**Svenska:** Professionell nivå

● 2024-05	<b>FABRIC ANALYTICS ENGINEER ASSOCIATE</b> Microsoft
● 2024-02	<b>AZURE DATA ENGINEER ASSOCIATE</b> Microsoft
● 2023-11	<b>AZURE DATA FUNDAMENTALS</b> Microsoft
● 2023-09	<b>AZURE AI FUNDAMENTALS</b> Microsoft
● 2023-08	<b>AZURE FUNDAMENTALS</b> Microsoft
● 2023-03	<b>HANDS ON ESSENTIALS - DATA ENGINEERING</b> Snowflake
● 2023-03	<b>HANDS ON ESSENTIALS - DATA LAKE</b> Snowflake
● 2022-11	<b>GETTING STARTED WITH AWS STORAGE</b> AWS Training Online
● 2022-04	<b>JINJA, MACROS, PACKAGES</b> DBT
● 2022-03	<b>DBT FUNDAMENTALS</b> DBT
● 2022-03	<b>SNOWFLAKE FUNDAMENTALS 4-DAY</b> Snowflake
● 2022-02	<b>HANDS ON ESSENTIALS - DATA SHARING</b> Snowflake
● 2022-01	<b>HANDS ON ESSENTIALS - DATA APPLICATIONS</b> Snowflake
● 2021-12	<b>HANDS ON ESSENTIALS - DATA WAREHOUSE</b> Snowflake
● 2016	<b>"ATT VARA EN LEDARE"</b> Bilda Sydöst, Högsby

# PROJEKT:

## Egen utvecklat verktyg för kvalitativa dataflöden (DIML):

Diml är ett XML-baserat verktyg som Region Skåne håller på att utveckla för att hantera kvalitativa dataflöden och säkerställa interoperabelt mellan olika system. Verktöget använder befintliga och öppna standarder så mycket som möjligt, som exempelvis DCAT-AP, OpenAPI och BPMN, för att underlätta integration och datautbyte.

Med hjälp av strukturell metadata gör det möjligt:

1. Minimera manuellt arbete genom att automatisera processer
2. Säkerställa leverantörsoberoende standarder
3. Bygga på metadata som grund för all datahantering, vilket underlättar framtida integrationer och analyser.

**Metod:** Agil, Scrum

## Datasjö för hälsodata:

Lagra och analysera vårddata från två huvudsakliga system inom Region Skåne för att möjliggöra maskininlärningsmodeller att förutsäga:

1. Sannolikheten för chock.
2. Sannolikheten för att en patient skrivs ut inom tre dagar.

På grund av datanskänslighet var arkitekturen främst on-prem till en lokal SQL Server.

Data extraherades och krypterades med hjälp av SSIS-paket, genererade med egenutvecklade verktyg.

Transformer och aggregering av data gjordes i SQL enligt Kimball-modellen.

Datakvalitet validerades och säkerställdes i Jupyter Notebook.

Tränade SVM-baserade ML-modeller on-prem.

Skapade interaktiva dashboards i Python tillsammans med en överläkare.

**Ramverk:** Agil, Scrum

## Utveckling och underhåll av Data Pipeline:

Lagras och analyseras telekomfinansiell data från olika team (fakturerings, kreditkontroll, försäljning, marknadsföring, ...) för att generera insikter på olika dimensioner som region, land, produkt. Genom att använda ELT-pipelines var de huvudsakliga stegen:

1. Sätta upp dagliga, månatliga eller timvisa procedurer med Luigi för att lagra data i en temporär AWS S3-bucket.
2. Flytta data till en permanent bucket i Parquet-format, med hjälp av Python.
3. Transformera data, slå samman med masterfiler och aggregera den, med hjälp av PySpark.
4. Validera och säkerställa datakvalitet med hjälp av Jupyter Notebook.
5. Lagra data i MySQL eller Snowflake DB.
6. Skapa dashboards i Tableau, baserat på intressenternas krav.

**Metod:** DevOps

## Migration projekt: MySQL -> Snowflake

En stor migrationsprocessen från MySQL till Snowflake. Omfattningen har varit att flytta alla tabeller, vyer och procedurer till Snowflake. Stegen har varit:

1. Granska den nuvarande MySQL-databasen och ta bort redundanser för att börja med.
2. Dokumentera alla Tableau-datakällor som behöver uppdateras för att peka på Snowflake-databasen.
3. Granska, dokumentera och skapa alla databaskonstverk som behöver återskapas i Snowflake.
4. Dokumentera alla pipelines i datalagret som behövde uppdateras för att peka på Snowflake-bucketen.
5. Uppdatera pipelines så att de också pekar på den nya Snowflake-bucketen.
6. Testa och validera data tillsammans med de olika dimensionerna.
7. Uppdatera pipelines för att ta bort MySQL-utgångsjobb efter att ha gått live med Snowflake

**Metod:** DevOps

## **Skapande och underhåll av Data Pipeline:**

Skapa och underhålla CI/CD-pipelines för att ta in data från olika källor (Excel, API:er, databaser, ...) till datalagret. Flytta sedan data mellan de olika AWS S3-bucketarna, med hjälp av Luigi som ett orkestreringsverktyg. De huvudsakliga lagren i ETL-processen var:

1. Första lagret: Validera data och dess integritet, ta sedan in och lagra rådata som den är.
2. Andra lagret: Transformera data och tillämpa affärslogik på den, validera den, och flytta sedan en kopia till en annan lagringsplats.
3. Tredje lagret: Exportera den transformerade datan till en Snowflake-databas.
4. Slutligen, skapa en Tableau-datakälla baserad på datan i databasen och skapa dashboards baserade på den.

**Metod:** DevOps

## **Språkagnostisk sentimentklassificerare för meddelanden**

Masteruppsatsprojekt genomfört på Sinch. Projektets mål är att begränsa hatretorik i dagens samhälle genom att använda en förtränad transformer-baserad modell för att identifiera sentimentet i ett textmeddelande och klassificera det som Positivt, Neutralt eller Negativt. Utmaningen var att producera en språkagnostisk modell som fungerar för vilket språk som helst, med hög klassificeringsnoggrannhet och låg inferenstid.

Vi jämförde ett dussin olika modeller med olika finjusteringstekniker för att uppfylla kraven. I slutet blev projektet en framgång och vårt arbete togs i produktion (och är fortfarande i produktion idag)

**Metod:** Agilt

