



## Ask Juhl Markestad

### Data Scientist / Data Engineer

Ask har en bakgrunn i teoretisk fysikk, hvor han fokuserte på matematisk modellering i fra Universitetet i Oslo. Han har erfaring som universitetslektor og har bidratt med artikler om fysikk i et populærvitenskapelig magasin. Han er entusiastisk når det gjelder dataanalyse, oppdagelsen av nye sammenhenger, kreativ problemløsning og viser alltid interesse for å utforske nye temaer og problemstillinger. For øyeblikket er han engasjert i utforskningen av maskinlæringsapplikasjoner, med særlig vekt på prosessen fra grunnleggende datagrunnlag og løsningsarkitektur til implementering og overvåking i produksjon, og med hvordan språkmodeller og generativ AI vil påvirke vår industri.

I sitt siste langvarige oppdrag har Ask utviklet et rammeverk for automatisert journalistikk og AI-komponenter tilknyttet dette. Han har også opplæring og sertifiseringer innen dataengineering og implementeringsteknikker, både for on-premise systemer og skybaserte løsninger. Ask er analytisk av natur, trives med utfordringer, og er dyktig til å identifisere muligheter og stille de rette spørsmålene. Han er også en fremragende formidler, som evner å forklare komplekse problemstillinger på en forståelig måte.

- ☐ Data Scientist
- ☐ Utvikler / Data Engineer

#### Data Science

LSTM, Natural Language Processing, XGBoost, Maskinlæring, Random Forest, Deep Learning, Statistics sampling, Databerikning, ML Testing, Logistisk Regresjon, Unsupervised Clustering, Support Vector Machines (SVM), Lineær Regresjon, Convolutional Neural Networks, Decision tree, Dataanalyse, Data Engineering, Mønsterkjennelse, Big data, Supervised learning, Unsupervised Learning, ML Produksjonspesifikasjoner, Vedlikehold av ML modeller, BERT, Transformers, Autoregression, Variational Autoregression, ChatGPT, Prompt Engineering, LLM, LLama, Retrieval Augmented Generation, RAG

#### Rammeverk

Scikit-learn, Pandas, Keras, Tensorflow, Numpy, PyTorch, Spacy, face\_recognition, tsfresh, statsmodel, pandas\_ta, PyTest, HuggingFace, FastAPI, Pydantic, SQLAlchemy, PySpark, Apache Spark, Vector database, OpenAI, LangChain, RAGA, RAG Triad of Metrics

#### Verktøy

Jupyter Notebook, Git, SQL, Bitbucket, Visual Studio Code, PyCharm, Postman, Poetry, TeamCity, Jenkins, flake8, black, mypy

#### Skyteknologi

AWS Sagemaker, Microsoft Azure, AWS, Azure Databricks, Azure Datafactory, Azure Machine Learning, Azure Virtual Machines, AWS StepFunctions, AWS Athena, AWS CLI, boto3, AWS CodeArtifact, AWS S3, AWS Lambda, AWS ECR, AWS CloudWatch, AWS BatchJobs, Databricks, Azure Synapse, Azure OpenAI, Azure AI Search

## Bransjer

---

Bank, finans og forsikring • Annet • Media • Shipping • Offentlig • HR, rekruttering og bemanning



# Prosjekterfaring

**02.2024 -****Policy Assistant for Fjordbank**

Fjordbank har, som alle banker, mye lovgivning og krav de må oppfylle. For å gjøre dette har de oppsamlet en stor mengde med policy dokumentasjon som i dag krever mye manuelt arbeid å jobbe med. Fjordbank er derfor i gang sammen med Miles i å utvikle en MVP for å lage en chatte applikasjon basert på Retrieval Augmented Generation (RAG). Prosjektet involverer å finne en passende løsning som er sikker og trygg nokk til å fungere for en bank, og som kan videre ekspanderes for fremtidige bruksmål.

**Rolle: Data Scientist**

Min rolle er som domene ekspert på LLMs og kunstig intelligens. Jeg er med for å hjelpe teamet å ta riktige teknologi valg, kvalitetsvurdere deres løsningsarkitektur, og å hjelpe utvikle løsningen. Vi vurderer flere mulige løsninger i fra bruk at betalte tjenester til å hoste løsningen selv med open source modeller. Vi ser spesielt på bruk av Azure OpenAI service sammen med Azure AI search kontra å bruke open source LLama 2 modeller for både embedding og chat, med Langchain som connector.

Som del av løsningsforslaget undersøker jeg hvordan å implementere gode evaluerings metoder og monitorering for sikre gode DevOps prosesser og enklere vedlikehold.

**Anvendt Kompetanse:**

LLama v2, Azure OpenAI, HuggingFace, Azure AI Search, LangChain, Retrieval Augmented Generation, RAGA, RAG Triad of Metrics

**04.2023 -****Pre-Sales for Miles**

I 2023 så Miles et behov for å skalere opp salgsarbeidet innenfor Data Science og Data Engineering. Ha flere kundemøter, flere workshops, flere ekspertoppdrag, for å skape flere muligheter. Som sentral person i Miles sin AI-satsning har jeg tatt mer ansvar i det strategiske arbeidet mot kunder, og for interne prosjekter innen fagfeltet. Jeg har og hatt ansvar for teknisk kompetanse i møte med kunder og for å sette meg inn i ny teknologi som kunder forespør slik som LLM og RAG.

**Rolle: Teknisk Ekspert Data Science og Data Engineering**

I denne rollen har jeg hatt ansvar for forberedelse til kundemøter, deltakelse i kundemøter, å holde workshops for kunder, gjøre kompetansehevende steg internt, og vært sparringspartner for interne ressurser og for kunder.

Oppgavene mine har handlet om å forstå kundenes problemer og kunne gi dem mulige løsninger. Arbeidet har omhandlet mange nye teknologier som ChatGPT, Large Language Models (LLMs), Retrieval Augmented Generation (RAG), og dataplattformteknologier slik som Databricks, etc. Jeg har laget løsningsarkitekturer rundt dataplattformer, og bruk av ChatGPT. I kundemøter har jeg bidratt med teknisk ekspertise og kunnet gi rådgivning som adresserer kundenes problemstillinger.

Som del av denne rollen har jeg holdt flere foredrag, både internt og eksternt, for å dele kompetanse. Spesielt eksternt har dette vært viktig for å generere flere salgsmuligheter for Miles.

**Anvendt Kompetanse:**

ChatGPT, LLM, Databricks, Azure Databricks, Azure Synapse, Microsoft Azure, OpenAI, Azure OpenAI, Løsningsarkitektur, Rådgivning, RAG, RAGA, RAG Triad of Metrics, Azure AI Search



09.2023 - 09.2023

Due Diligence for Idekapiital

Jeg ble hyret til å gjøre en teknisk due diligence av et selskap innen AI sektoren som Idekapiital vurderte å investere i. Jeg vurderte om produktet løste et reelt problem, om produktet løser dette problemet, om det løser problemet bedre enn konkurrenter i markedet, hvor vanskelig det er for konkurrenter å replikere produktets funksjonalitet, hvilke kundesegmenter er interessert i et slikt produkt, og hvor stor betalingsvillighet det er i markedet for et slikt produkt.

Anvendt **Kompetanse:**

Due diligence, Markedsanalyse, Teknisk analyse

08.2023 - 09.2023

dobee.ai for dobee

dobee er et OKR orkistreringsverktøy som hjelper kunder å lage gode mål og resultater, å formilde arbeidet, og gi dem evnen til å ha god oversikt over deres OKR reise. For å bistå sinde kunder har dobee lansert sitt dobee.ai prosjekt som lager gode mål og resultat forslag basert på kundens informasjon, og hjelper kunden komme i gang med å bruke OKR.

dobee trengte hjelp med å bekrefte deres prosjekt planer og vurdere beste fremgangsmåte for å nå målet. Samtidig trengte dobee hjelp med å oversatt domene ekspertisen deres til bruk i dobee.ai. Videre ønsket dobee en vurdering av deres egne data og hvilken grad de kunne bruke dem til sitt AI prosjekt.

Rolle: **Data Engineer**

Jeg tokk en rådgivende rolle der jeg gjorde en vurdering av deres konsept, lagde en arkitektur skisse for målbilde og en detaljert roadmap for å nå målet. I løpet av denne prosessen gjorde jeg intervju for å samle inn domene kunnskap om OKR for å vurdere hvilke tiltak som gir høyest verdi og får å sette prioriteringer i roadmap. For å skape noe umiddelbar verdi oversatte jeg noe av denne domene kunnskapen in i dobee.ai som er ChatGPT basert.

Videre så jeg på deres datagrunnlag og deres fremtidige planer for tjenseter og til hvilken grad disse vil gi mer verdi for AI prosjektet.

Anvendt **Kompetanse:**

ChatGPT, Prompt Engineering, Arkitektur, Road map, Vector database, SQL

01.2022 - 12.2022

Automatisert Journalistikk for NTB

Våren 2022 startet NTB opp en ny avdeling med mål om å sette en bransjestandard innen automatisert journalistikk. De ønsket å frigjøre tid for journalistene ved å la algoritmer skrive rutinesakene for dem, bygge et robot-rammeverk som kunne brukes av journalister for å lage hyperlokale vinklinger om det som skjer i distriktene, forenkle oppgaven for redaksjonelle utviklere når nye roboter skal på plass samt å bygge en api-portal for å distribuere automatiserte artikler ut til alle Norges største mediehus. For å nå disse målene ville NTB ta i bruk Natural Language Processing og cutting edge AI for å lage et rammeverk for å produsere autmatisk genererte journalistiske verk. I tillegg skulle det gjøres et sky-løft der all dagens infrastruktur skulle flyttes fra Openshift til AWS serverless.

<https://www.ntb.no/nyheter/automatisert-journalistikk>

<https://m24.no/automatisering-automatisert-journalistikk-lasse-lervik/ntb-opprettet-egen-avdeling-for-automatisert-journalistikk/460590>

Min oppgave i dette prosjektet var å hjelpe dem i gang med å bygge rammeverket, lage datapielines, hjelpe dem å finne en data scientist for fast ansettelse, å begynne å se på og bygge AI komponenter til et slikt rammeverk.

## Rolle: **Utvikler / Data Engineer**

På grunn av en umodenhet i den eksisterende MVP på prosjektet var det behov for å utvikle et solid rammeverk i bunn før det var vits i å se på mulige AI komponenter. De første 9 månedene ble brukt med utvikling av dette rammeverk med fokus på følgende tjenester og løsninger. NTB gikk over til en mikrotjeneste arkitektur i AWS underveis i prosjektet. Hver tjeneste var et docker image, som ble sendt til TeamCity og videre til Octopus som orkistrerte deployment i AWS.

### Dataingesttjeneste:

En tjeneste for parsing av data fra dataleverandør, strukturering av denne dataen i en Pandas dataframe, og rengjøring av dataen for videre prosessering.

### Datatransformasjontjeneste:

En tjeneste for å lage nye feature, transformere og formatere disse dataene til å være tekstklare. For dette formål ble et generelt rammeverk lagd som kalte på forskjellige pre-definerte transformasjonsfunksjoner, som kunne konfigureres fra et eksternt json dokument.

### Designet template løsningen:

Et schema for teksttre som brukes som templates for å sette sammen en tekst. Denne template løsningen måtte kunne konfigureres med flere typer tester av datainnhold, styre rekkefølgen på teksten i produsert artikkel, og sette kontroller på datakvalitet.

### Artiklerskrivertjeneste:

En tilsvarende parser og skriver for å produsere teksten konfigurert i templateløsning, samt fungere som integrasjonsflate for fremtidige AI tjenester.

### Standardisering:

Jobbet med eksisterende team for å etablere en standardisert representasjon av nyheter i JSON (NinJS - IPTC). Vi satte opp en postgres database for å serve denne standarden der vi brukte Pydantic for å sette opp database, sette schema krav, og brukte Pydantic sammen med FastAPI for å lage et REST API for kundene å bruke.

## Rolle: **Data Scientist**

Som data scientist jobbet konsulenten med å planlegge AI komponenter, ansettelse av data scientist for å ta over konsulentens rolle, og bygging av AI komponenter. I dette prosjektet jobbet konsulenten sammen med den ansatte data scientisten til å bygge en synonym generator for å automatisk kunne skrive om enkelt ord i en tekst for å lage mer variasjon mellom tekster produsert.

Dette gjøres ved å bruke en BERT masking model for å lage synonymforslag, som blir ord for ord sammenlignet med originalordet og de beste blir vurdert videre. Vi kjører da en sentence BERT model for å sjekke likhet mellom original og produsert tekst. Dermed får vi en sjekk på at den produserte teksten har samme betydning som originalen. Videre blir synonymene sammenlignet med synonymreferansedataset fra nasjonalbiblioteket for å sikre eventuelle feil fra modellene.

I tillegg lagdes en 'harmoniserer' som brukte Spacy til å analysere en tekst for avhengigheter mellom ord, spesifikt for adjektiver, finne hvilket bøyingsform disse ordene burde ha og automatisk fikse feilbøyninger. Dette ble brukt for å sikre at tekst blir grammatisk riktig etter en automatisk omskriving som ved synonymgeneratoren beskrevet over.

## Anvendt **Kompetanse:**

Python, PyTest, BERT, Spacy, Pandas, Transformers, HuggingFace, PyTorch, FastAPI, Pydantic, Poetry, SQLAlchemy, Docker, AWS CodeArtifact, Natural Language Processing, Lucidchart, Bitbucket, Visual Studio Code, PyCharm, boto3, AWS CLI, Postman, TeamCity, AWS ECR, AWS CloudWatch, flake8, AWS S3, black, Arkitektur

08.2020 - 12.2021

Data Scientist for Western Bulk

Data scientist - Prosjektet utarbeidet modeller som skulle predikere markedet og agere på markedsutviklingen med automatiserte modeller. Designe og utvikle metoder og produkter for analyse og prosessering av data. Bygge modeller for å prognosere markedet, og bygge modeller for å gi anbefalinger og handlinger basert på disse prognosene, primært innen for trading og risikovurdering.

## Rolle: Data Scientist

Konsulenten jobbet med å utvikle verktøyene og bygge modellene for markedsprediksjonene. I løpet av prosjektet var konsulenten med på å designe og bygge metoder for å prosessere og formatere tidsseriedata for bruk i markedsprognosene. For å kunne bygge relevante og "up to date" modeller måtte det designes og bygges et automatisert dataanalyse og feature-selection verktøy, der konsulenten hadde hovedansvar for feature-selection verktøyet. Videre jobbet konsulenten med å designe og bygge et sekvensielt trene- og testrammeverk for prognosemodeller. Konsulenten jobbet også med å designe, lage, og hovedansvarlig for å tune prognose sub- og ensamble modeller. Konsulenten var også involvert i å designe de automatiserte modellene som skulle agere på prognosen. Disse modellene handlet primært om trading-strategier og risikovurdering.

For monitorering av innkommende data og modell ytelse brukte teamet flere custom dashboards laget i PowerBI, men med egne applikasjoner i bunn som gjorde alle data prosessering stegene. Dermed ble PowerBI kun brukt for å lage en GUI for å presentere data, ikke for å gjøre utregninger. Konsulenten bygget flertallet av disse applikasjonene, som leverte csv-filer til PowerBI.

Hele kodebasen var bygget i Python og hostet i AWS. Alle komponenter var bygget av teamet, gjennom kjente data science pakker, og mer nisje pakker for tidsserier som tsfresh og pandas\_ta. Komponenter ble kjøpt med AWS Lambdas eller AWS Batch Jobs i stacks konfigurert gjennom AWS Step-Functions.

SQL ble brukt for å generere tabeller gjennom AWS Athena på partisjonert fil system, for videre analyse med Python verktøy i notebooks og for bruk i produksjons-pipeline.

## Anvendt Kompetanse:

AWS, tsfresh, statsmodel, pandas\_ta, Produksjonssetting, Monitoring, Vedlikehold av ML modeller, Trading Modeller, Risikoanalyse, AWS StepFunctions, SQL, AWS BatchJobs, AWS Lambda, Jenkins, flake8, black, mypy, Autoregression, Variational Autoregression

04.2019 - 08.2019

Forprosjekt-Omsyn for OBOS

Undersøke om Omsyn konseptet passer for OBOS og lage et løsningsforslag til implantasjon. Omsyn konseptet handler om å kunne gi automatisert personlig rådgivning ved å bruke kunstig intelligens til å klassifisere transaksjoner som brukes sammen med CRM data for å lage bedre kundesegmentering som treffer på vanene dine fremfor tradisjonell kundesegmentering som aldersgruppe. Denne kundesegmenteringen vil brukes for å gi tips og råd basert på din faktiske situasjon samt kunne gi tilbud fra OBOS partnere som faktisk ønskelige for medlemmene.

## Rolle: Data Scientist

Gjennomgang av datagrunnlaget til OBOS, datainnhentings mulighetene deres og vurdering av mulighetsrom for prediksjon av livsfase til medlemmene, samt kartlegge implementasjon av løsningen.

11.2018 - 07.2019

Data Analyse for Cubit AS

Cubit AS er en start-up som har utviklet en plattform for å registrere EL sikkerhetsinspeksjoner. Ask har gjort en analyse av Cubit AS samlede data og utforskning av mulighetsområde for bruk av maskinlæring innen risikoanalyse.

## Rolle: **Data Scientist**

Behandlet data ved rengjøring av ufullstendig data, og transformering dato og tekst beskrivelse til et maskinleselig input. Brukte den behandlede data til å predikere EL sikkerhetsfeil i boliger.

## Anvendt **Kompetanse:**

Logistisk Regresjon, Pandas, Scikit-learn, Supervised learning

05.2019 - 06.2019

Utforskning av data grunnlag og maskinlærings muligheter for Kommuneforlaget

AS

Kommuneforlaget drifter spørreundersøkelser om kvalitative opplevelser for offentlige ansatte i kommuner, for innbyggere og brukere av offentlige tjenester. De har dermed en unik posisjon ved å samle kvalitativ data og vil utforske kvaliteten og bredden på data grunnlaget deres samt utforske hvilke muligheter som eksisterer for anvendelser av maskinlæring med eksterne datakilder.

## Rolle: **Data Scientist**

Samlet inn eksterne kilder og etablerte en oversikt over dem. Formaterte data fra flere ulike kilder og kombinerte dem for å generere datasett for utforskning av sammenhenger og hypotese testing ved å anvende maskinlærings modeller. Se på fremtids muligheter med maskinlæring for Kommuneforlaget og hvordan ta en unik rolle fremover i et voksende marked for maskinlæring og IoT.

## Anvendt **Kompetanse:**

Statistics sampling, Pandas, Data Engineering, Scikit-learn, Jupyter Notebook, Logistisk Regresjon

03.2019 - 03.2019

Kvalitets vurdering av arkitektur tilbud og forslag for arkitektur for Culturengine

Kvalitets vurdering av et arkitektur forslag for speed-up av eksisterende prosesser med mulighet for fremtidig applikasjon av maskinlæring i Amazon Web Services, samt eget forslag for modul basert arkitektur. I tillegg gjorde vi en vurdering av deres datagrunnlag, muligheter fremover med maskinlæring, og når datamengden vil være tilstrekkelig for å begynne arbeidet med maskinlæring.

## Rolle: **Data Scientist**

Utforskning av eksisterende datagrunnlag, vurdering av data innsamlings prosessene, vurdere mulige applikasjoner av maskinlæring, og evaluering AWS sine dedikerte maskinlærings moduler.

## Anvendt **Kompetanse:**

Prosjektevaluering

11.2018 - 02.2019

Forarbeid Omsyn for Miles AS



Utforske tilnærming til bruk av maskinlæring for analyse og klassifisering av transaksjons data, og videre analyse av klassifiseringene for innsikt givende presentasjon ovenfor bankkunder.

Transaksjons data inneholder mye informasjon som i dag ikke blir brukt. Målet med prosjektet er å ekstrahere denne ubrukte informasjonen ved å identifisere hvilken kategori en transaksjon tilhører, gjengi informasjonen til sluttkunden og bruke informasjonen sammen med kundeopplysninger til å lage bedre, mer personlige kundesegmenter slik at banken kan tilby personlige råd automatisert gjennom bankens egne digital kundeflater.

## Rolle: Data Scientist

Analyse av transaksjons klassifiserings problemets kompleksitet; tilgjengelige variabler, kvalitet på data, lineære vs ikke-lineære avhengigheter, muligheter for data beriking med eksterne kilder, optimaliserings metrikker, monitoring i produksjon, og muligheter for vedlikehold av modellen. Videre valg av type modell, muligheter for feature engineering, og feature extraction.

Sett på maskinlærings metoder for kundesegmentering og muligheter for kombinasjon med tradisjonell buisness logikk.

## Anvendt Kompetanse:

Databerikning, Prosjektplanlegging, Natural Language Processing, Evaluering og optimalisering av modeller

## Utdannelse

2015 - 2017	Universitetet i Oslo	Master i teoretisk fysikk
2012 - 2015	Universitetet i Oslo	Bachelor i Fysikk

## Sertifiseringer

2023	<b>Databricks Certified Associate Developer for Apache Spark 3.0</b> Databricks
2020	<b>Microsoft Certified: Azure Data Scientist Associate</b> Microsoft
2020	<b>Machine Learning Engineer Nanodegree</b> Udacity
2019	<b>Deep Learning Specialization</b> Coursera
2019	<b>SQL for datascience</b> Coursera
2019	<b>TensorFlow in Practice</b> Coursera





## Kurs

2023	<b>Partner Solutions Architect Essentials (Champion Series)</b> Databricks
2023	<b>Apache Spark Developer Learning Plan</b> Databricks
2020	<b>DP-201: Designing an Azure Data Solution</b> Glasspaper
2020	<b>DP-200: Implementing an Azure Data Solution</b> Glasspaper
2019	<b>Deployment of Machine Learning Models</b> Udemy
2019	<b>The Analytic Edge</b> edX-MIT
2018	<b>Machine Learning Foundations</b> Udacity

## Kompetanse

### Data Science

LSTM, Natural Language Processing, XGBoost, Maskinl ring, Random Forest, Deep Learning, Statistics sampling, Databerikning, ML Testing, Logistisk Regresjon, Unsupervised Clustering, Support Vector Machines (SVM), Line r Regresjon, Convolutional Neural Networks, Decision tree, Dataanalyse, Data Engineering, M nstergjenkjenning, Big data, Supervised learning, Unsupervised Learning, ML Produksjonspesifikasjoner, Vedlikehold av ML modeller, BERT, Transformers, Autoregression, Variational Autoregression, ChatGPT, Prompt Engineering, LLM, LLama, Retrieval Augmented Generation, RAG

### Rammeverk

Scikit-learn, Pandas, Keras, Tensorflow, Numpy, PyTorch, Spacy, face\_recognition, tsfresh, statsmodel, pandas\_ta, PyTest, HuggingFace, FastAPI, Pydantic, Sqlalchemy, PySpark, Apache Spark, Vector database, OpenAI, LangChain, RAGA, RAG Triad of Metrics

### Verkt y

Jupyter Notebook, Git, SQL, Bitbucket, Visual Studio Code, PyCharm, Postman, Poetry, TeamCity, Jenkins, flake8, black, mypy

### Skyteknologi

AWS Sagemaker, Microsoft Azure, AWS, Azure Databricks, Azure Datafactory, Azure Machine Learning, Azure Virtual Machines, AWS StepFunctions, AWS Athena, AWS CLI, boto3, AWS



CodeArtifact, AWS S3, AWS Lambda, AWS ECR, AWS CloudWatch, AWS BatchJobs, Databricks, Azure Synapse, Azure OpenAI, Azure AI Search

## Metodikk

Matematisk modellering, Behovskartlegging / kravspesifisering, Lucidchart, Evaluering og optimalisering av modeller, CI/CD, Produksjonssetting, Monitoring, Trading Modeller, Risiko Modellering, Road map, Due diligence, Teknisk analyse

## Språk/plattform

Python, R, C++

## Infrastruktur

Docker

## Arkitektur

Arkitektur, Løsningsarkitektur

## Arbeidserfaring

2018 -	Miles Oslo AS	Data Scientist
2017	Universitet i Oslo	Universitetslektor
2016	Universitetet i Oslo	Gruppelærer

## Språk

Norsk Morsmål

Engelsk Flytende

