**Business Intelligence og dataanalyse til data-drevne beslutninger i GAME ngo**

Hovedopgave – Pelle Hald Vedsmand

Contents

[1. Forord 1](#_Toc183088837)

[2. Indledning 1](#_Toc183088838)

[3. Virksomhed 2](#_Toc183088839)

[4. Krav 3](#_Toc183088840)

[5. Valg af teknologi(er) 5](#_Toc183088841)

[6. Valg af arkitektur 5](#_Toc183088842)

[7. Design 6](#_Toc183088843)

[8. Implementering 6](#_Toc183088844)

[9. Test 6](#_Toc183088845)

[10. Evaluering / konklusion 6](#_Toc183088846)

# Forord

Efter mit praktikophold hos GAME i perioden 5/8 – 28-11, har jeg været i samarbejde om at udvikle et produkt for firmaet.  
**Udvikling i praktikophold**:  
Under min praktikperiode skabte jeg en pipeline, der skulle automatisere og forbedre GAMEs arbejde med dataanalyser. Derudover hjalp jeg forskellige teams i GAME med analyse og modellering af data, dog på projektbasis.

# Indledning

Hvordan kan data-drevne beslutninger skabe success for GAME? Og hvordan kan data engineering samt Business Intelligence bruges til at skabe disse beslutninger?  
Dette var den overordnede problemstilling i virksomheden.   
Denne rapport vil beskrive, hvordan man kan bruge Business Intelligences principper og værktøjer til at skabe analyser og modeller, som vil kunne påvirke GAMEs vækst som firma.

Under min praktikperiode håndterede jeg dataanalysearbejde i flere projekter. Heri fik GAME en forståelse for mulighederne med Business Intelligence.  
  
I følgende rapport er der opstillet forskellige prediction models, baseret på data i og udenfor GAME, der skal gøre det muligt at forudsige deltagelsen til fremtidige events.  
GAME vil med denne viden kunne optimere events med flere/færre instruktører og materiel. De vil også få dannet en større forståelse for diverse deltagelsesmønstre, samt korrelationen mellem deltagelsestal og features, både i GAME, men også udefrakommende.

# Virksomhed

GAME er en ngo der forsøger at skabe social forandring hos børn og unge igennem aktiviteter og sport. GAME træner unge frivillige til at være instruktører og rollemodeller[[1]](#footnote-1)

GAME er en NGO, der er sponsoreret af diverse fonde, firmaer og legater. GAME skal derfor ud at bevise for investorerne, at de har formået at løse de opgaver der stilles.

Udviklingen i GAME er meget flydende i GAME, dvs. at de ikke følger nogle stringente modeller eller metoder. Det meste af arbejdet foregår på projektbasis, gennem investorer.  
Disse projekter ledes af flere forskellige teams, der alle har brug for analyser af indkommende data.

**Formål:**

Udviklingen af prediction models vil hjælpe GAME med at vokse som virksomhed, da de vil kunne skabe bedre finansielt råderum, ved at kunne distribuere budgettet til de rigtige events.

Forståelse af sammenhængen mellem datamønstre i og udenfor GAME, og hvordan de kan forbedre deltagelsen til fremtidige events.

**Problemstillinger:**

* GAME har problemer med at forudsige antal deltagere til fremtidige events, og de bruger derfor, til tider, for mange penge på events med lav deltagelse.
* GAME har svært ved at forstå hvilke ”features”, der påvirker deres deltagelsestal til events.
* Hvordan kan data hjælpe med at forudsige deltagelsen i GAME husene i Danmark
* Hvordan kan data bruges til at forbedre deltagelsen i GAME husene?

**Mål og resultater**

Projektets formål er at hjælpe GAME med at blive en mere data-drevet organisation, der kan udnytte dataanalyse og Business Intelligence til at opnå konkrete forbedringer i deres aktiviteter og ressourcestyring. Dette omfatter følgende mål:

**Økonomiske mål**

* Reducere unødvendige omkostninger ved events med lav deltagelse ved hjælp af præcise forudsigelser af deltagertal.
* Øge GAMEs finansielle råderum ved mere effektiv allokering af ressourcer og budget.

**Operationelle mål**

* Forbedre effektiviteten i planlægning og afholdelse af events gennem automatiserede dataanalyser og prediction models.
* Skabe hurtigere beslutningsprocesser baseret på indsigter fra data.

**Kundemæssige mål**

* Øge kundetilfredshed ved at optimere antallet af instruktører og materialer til events, så deltagerne får en bedre oplevelse.
* Forbedre deltagelsesmønstrene ved at identificere og handle på de vigtigste faktorer, der påvirker deltagelsen.

**INVEST-mål**

Målene er defineret efter INVEST-principperne:

* **Independent**: Modellerne kan implementeres og bruges uden afhængighed af andre systemer.
* **Negotiable**: Målene er fleksible og kan tilpasses GAMEs specifikke behov.
* **Valuable**: Modellerne skal skabe værdi ved at reducere omkostninger og forbedre kundetilfredshed.
* **Estimable**: Succeskriterierne for modellerne kan vurderes gennem målbare resultater som f.eks. forbedret nøjagtighed i forudsigelser.
* **Small**: Målene er opdelt i små, overskuelige opgaver, der sikrer en agil udviklingsproces.
* **Testable**: Modellerne testes for at sikre deres funktionalitet og relevans for GAMEs behov.

Disse mål sikrer, at projektet bidrager til GAMEs overordnede mission om at skabe social forandring gennem sport og aktiviteter, samtidig med at der skabes konkrete forbedringer i organisationens drift.

# Krav

**Omfanget af produktet (Scope)**

I dette projekt er formålet at udvikle en datadrevet applikation, som hjælper GAME med at forbedre planlægning og gennemførsel af events. For at sikre en klar forståelse af, hvad applikationen skal kunne, er kravene beskrevet gennem **user stories** med tilhørende **acceptance criteria**. Disse krav sikrer, at applikationen opfylder både funktionelle og non-funktionelle behov hos GAME.

**Funktionelle krav**

**User Story 1: Forudsigelse af deltagertal for Playmaker Teamet**

* **Beskrivelse:** Som et teammedlem hos playmakerteamet vil jeg kunne forudsige deltagertal for fremtidige events, så jeg kan allokere de rette ressourcer (instruktører og materialer) effektivt.
  + **Acceptance Criteria:**
    1. En model skal kunne estimere deltagertal baseret på historiske data og eksterne data (fx vejrdata eller helligdage).
    2. Output skal præsenteres i en brugervenlig Streamlit-applikation.
  + **Omfang:** Large – Flere modeller udvikles baseret på GAMEs egne data samt eksterne faktorer.

**User Story 2: Identifikation af vigtige faktorer**

* **Beskrivelse:** Som et teammedlem hos playmakerteamet vil jeg identificere de vigtigste faktorer, der påvirker deltagelse i events, så jeg kan give anbefalinger til optimering af fremtidige arrangementer.
  + **Acceptance Criteria:**
    1. En liste over de top 5 vigtigste faktorer (fx vejr, ugedag, eller eventtype) skal præsenteres i form af visualiseringer.
  + **Omfang:** Medium – Fokus på dataforberedelse og brug af modeller som Random Forest til feature importance-analyse.

**User Story 3: Analyse af historiske data**

* **Beskrivelse:** Som en analytiker hos GAME vil jeg kunne analysere historiske deltagelsesmønstre, så jeg kan finde tilbagevendende tendenser, der kan bruges til planlægning.
  + **Acceptance Criteria:**
    1. Analyse af deltagelsesmønstre skal inkludere visualiseringer af sæsonvariationer og trends.
    2. Resultaterne skal præsenteres i et format, der er let forståeligt for ikke-tekniske brugere.
  + **Omfang:** Small – Fokus på eksisterende data og basisanalyse.

**Non-funktionelle krav**

De non-funktionelle krav sikrer, at applikationen er effektiv og brugervenlig:

* **Ydelse:** Applikationen skal kunne levere forudsigelser hurtigt (under 5 sekunder) ved indtastning af nye data.
* **Skalerbarhed:** Designet skal være fleksibelt og kunne tilpasses, hvis nye funktioner eller datakilder tilføjes i fremtiden.
* **Brugervenlighed:** Streamlit-applikationen skal være intuitiv og kræve minimal træning for brugerne.
* **Sikkerhed:** Adgang til data og applikationen skal være begrænset til relevante medarbejdere i GAME.

# Valg af teknologi(er)

**Overvejelser og behov**

Valget af teknologi blev foretaget med fokus på at opfylde både virksomhedens behov og de krav, der var til projektet. GAME ønskede en løsning, der kunne levere praktisk værdi ved at analysere og præsentere data, samtidig med at den var intuitiv og nem at bruge. Mit mål var at udvikle en løsning, der ikke blot matchede GAMEs behov, men som også gav mig mulighed for at udforske og styrke mine færdigheder inden for dataanalyse og machine learning. Det var desuden afgørende, at teknologivalget tog højde for den begrænsede tid til projektet.

**Alternativer og deres udfordringer**

En af de løsninger, jeg overvejede, var at bygge en applikation med en Java-backend til databasestyring, Python til modellering og en React-baseret frontend. Denne arkitektur ville have været fleksibel og skalerbar, men den var for kompleks og tidskrævende i forhold til projektets omfang. Alternativt overvejede jeg Flask som en lettere backend-løsning kombineret med en dedikeret frontend. Flask kunne have leveret en enkel og tilpasselig webapplikation, men det ville stadig kræve en vis mængde tid og ressourcer til at udvikle og forbinde de forskellige dele.

**Endeligt valg: Streamlit**

Jeg valgte at bruge Streamlit som det primære framework til at præsentere resultaterne. Streamlit gjorde det muligt hurtigt at udvikle en løsning, hvor dataanalyser og visualiseringer kunne præsenteres på en brugervenlig måde. Det var et ideelt valg, fordi det fjernede behovet for omfattende frontend-udvikling og gjorde det muligt at fokusere på den centrale opgave: at udvikle og optimere datamodeller. Sammen med Python som kerneværktøj og biblioteker som Scikit-learn og TensorFlow var det muligt at bygge modeller til regression, klassifikation og clustering. Denne kombination sikrede, at jeg kunne levere en funktionel prototype, der opfyldte GAMEs behov inden for tidsrammen.

# Valg af arkitektur

**Overvejelser om arkitektur**

Arkitekturen til projektet blev designet med fokus på enkelhed og effektivitet. GAME havde behov for en løsning, der hurtigt kunne implementeres og demonstrere værdien af datadrevet beslutningstagning. På grund af den begrænsede tid og projektets primære fokus på dataanalyse blev det nødvendigt at vælge en arkitektur, der reducerede kompleksiteten og maksimerede hastigheden i udviklingen.

**Forskellige alternativer**

Jeg overvejede en flerlaget arkitektur, hvor backend og frontend blev udviklet separat. En mulighed var at anvende Flask som backend til at køre modellerne og håndtere dataintegration, mens en frontend i React kunne bruges til præsentation. Denne løsning ville have givet større fleksibilitet, men den krævede samtidig mere tid og flere ressourcer.

En anden overvejelse var at bruge en kombination af en ekstern database med et traditionelt REST API. Selvom dette ville have været en robust løsning, var det ikke nødvendigt for projektet, da datamængden og kravene til opbevaring ikke var særlig store.

**Streamlit som løsning**

Streamlit blev valgt som den samlede arkitektur til projektet, da det forenede frontend og backend i én løsning. Dette gjorde det muligt at fokusere på udvikling af datamodeller og samtidig skabe en interaktiv applikation uden at skulle bygge separate systemer. Streamlit var ideelt til hurtigt at præsentere de forskellige forudsigelsesmodeller og visualiseringer, som projektet krævede.

Ved at bruge Python sammen med Streamlit kunne jeg holde arkitekturen enkel og samtidig levere en løsning, der var skræddersyet til GAMEs behov. Arkitekturen gjorde det muligt at opnå en høj grad af funktionalitet, samtidig med at den var nem at vedligeholde og forstå for virksomhedens medarbejdere.

# Design

# Implementering

# Test

# Evaluering / konklusion

1. <https://game.ngo/what-we-do/> [↑](#footnote-ref-1)