**CRISP-DM**

**Business Understanding:**

**Huidige situatie**  
Klanten voeren prognoses van toekomstige bestellingen in binnen het 4FOOD|ERP-systeem In de praktijk wijken realisaties echter regelmatig af van deze prognoses, wat leidt tot voorraadproblemen en inefficiënte planningen. Tot op heden worden deze afwijkingen niet systematisch geanalyseerd of voorspeld.

**Businessdoelen**

* Estrategy wil kunnen vaststellen *wanneer* en *waarom* prognoses structureel afwijken van de realisaties.
* Door patronen te identificeren kan Estrategy klanten ondersteunen bij het verbeteren van hun voorspellingen.
* Een data-gedreven aanpak moet leiden tot verbeterde voorraadnauwkeurigheid en efficiëntere bedrijfsvoering bij klanten, inclusief verbeteringen in het 4FOOD|ERP-systeem.

**Dataminingdoelen**

* Systematisch inzichtelijk maken van afwijkingen (absolute en procentuele afwijking).
* Identificeren van patronen (klant, productgroep, seizoensinvloed).
* Bouwen van een voorspellend classificatiemodel dat kan aangeven of een nieuwe prognose waarschijnlijk correct, te hoog of te laag zal zijn.

**Success Criteria**

* De resultaten worden herkend en gevalideerd door interne experts en sluiten aan bij de praktijkervaring van klanten.
* De analyse is duidelijk communiceerbaar en kan worden toegepast in toekomstige 4FOOD|ERP-systeem-modules.
* Het classificatiemodel behaalt een acceptabele nauwkeurigheid (≥70% overall accuracy en voldoende F1-score per klasse).

**Data understanding:**

**Doel:** Het doel van deze fase is om een zo volledig mogelijk beeld te krijgen van de beschikbare data en de eigenschappen ervan. Dit betekent dat ik niet alleen data verzamel, maar ook leer begrijpen wat de data vertegenwoordigt, hoe deze is opgebouwd, en welke kwaliteit de data heeft.

* In de context van 4FOOD|ERP-systeem is het doel om inzicht te krijgen in hoe prognoses en realisaties zijn opgeslagen, welke variabelen beschikbaar zijn (klant, productgroep, periode, hoeveelheid), en in welke vorm de data in het systeem wordt bijgehouden.
* Daarnaast is het doel om relaties en patronen in de data te ontdekken, bijvoorbeeld seizoensinvloeden, structurele overschatting door klanten of afwijkingen per productgroep.
* Cruciaal onderdeel is de datakwaliteit: ontbrekende waarden en inconsistenties worden gedetecteerd en gedocumenteerd, omdat zij direct van invloed zijn op de betrouwbaarheid van de analyse en het model.
* Exploratory Data Analysis (EDA) opnemen: histogrammen, boxplots, correlaties.
* Uiteindelijk vormt deze fase de basis om te bepalen of de data geschikt is om de businessdoelen te beantwoorden, of dat aanvullende bronnen nodig zijn.

**Risico’s:**

* Onvoldoende datahistorie bij sommige klanten.
* Mogelijke vertekening door uitzonderlijke gebeurtenissen (bv. COVID, feestdagen).

**Data preparation:**

**Doel:** Het doel van deze fase is om ruwe data om te zetten naar een betrouwbare dataset. Het doel is om fouten en ontbrekende waarden te corrigeren, relevante data te selecteren en de dataset te verrijken zodat deze direct bruikbaar is voor analyse en modellering.

* In de context van 4FOOD|ERP-systeem betekent dit dat we prognose- en realisatiedata opschonen, transformeren en verrijken met contextuele variabelen.
* Het doel is dat alle data eenduidig, schoon en consistent is, zodat er geen ruis of fouten ontstaan in de analyses en modellen.
* Naast het opschonen wordt data verrijkt en voorbereid: zoals berekenen van afwijkingen (absoluut en procentueel), het toevoegen van seizoens-variabelen en het creëren van klant- of productgroep indicatoren.
* Klassenbalans aanpakken op training set (oversampling/undersampling).
* Ten slotte is het doel om de data te structureren voor de modelleerstap: consistente formaten, juiste splitsing in train- en testsets, en voldoende representatie van alle klassen (correct, te hoog, te laag).
* Deze fase is de meest arbeidsintensieve, maar vormt het fundament waarop alle verdere resultaten rusten.

**Modeling:**

**Doel:** Het doel van deze fase is om met de voorbereide datavoorspelmodellen en analysetechnieken te ontwikkelen en te testen, die de afwijkingen tussen prognose en realisatie kunnen verklaren en voorspellen.

* In de context van 4FOOD|ERP-systeem betekent dit het bouwen van een classificatiemodel dat kan voorspellen of een nieuwe prognose waarschijnlijk correct, te hoog of te laag zal zijn.
* Het doel is om verschillende algoritmes (zoals Random Forest, Gradient Boosting enz.) te vergelijken en te bepalen welk model de beste balans biedt tussen nauwkeurigheid, interpretatie en praktische toepasbaarheid.
* Een ander doel is om inzicht te krijgen in de belangrijkste variabelen die bijdragen aan afwijkingen, zodat er concrete feedback naar klanten gegeven kan worden (bijv. “Productgroep X wordt structureel overschat rond feestdagen”).
* Hyperparameter tuning opnemen.
* Explainability toevoegen (feature importance, SHAP).
* Omdat sommige modellen specifieke datavormen vereisen, kan het doel ook zijn om iteratief terug te koppelen naar de Data Preparation-fase, om extra features te genereren of transformaties door te voeren.

**Evaluation:**

Het doel van deze fase is om te beoordelen of de resultaten van de modellering daadwerkelijk voldoen aan de vooraf vastgestelde businessdoelen en dataminingdoelen.

* Binnen 4FOOD|ERP-systeem betekent dit toetsen of de modellen niet alleen statistisch goed presteren (≥70% accuracy en goede F1-scores), maar ook praktisch bruikbaar zijn voor klanten en planners.
* Gebruik van confusion matrix en precision/recall per klasse.
* Kostenmatrix meenemen (impact foutieve voorspellingen).
* Het doel is om de resultaten te laten valideren door domeinexperts: herkennen zij de patronen, en sluiten de inzichten aan bij hun praktijkervaring?
* Daarnaast is het doel om te bepalen of het model generaliseerbaar is: presteert het ook goed op nieuwe, recente data, en blijft het robuust over tijd?
* In deze fase wordt beslist of de oplossing gereed is voor implementatie in 4FOOD|ERP-systeem of dat er nog verbeteringen nodig zijn in eerdere stappen (bijv. meer features toevoegen, model aanpassen). (Zal de eerst paar iteraties niet van toepassing zijn.)

**Deployment:**

**Doel:** Het doel van deze fase is om de resultaten en modellen om te zetten in praktisch toepasbare oplossingen die waarde opleveren voor de organisatie en hun klanten.

* Voor 4FOOD|ERP-systeem betekent dit dat de inzichten niet alleen in een rapport gepresenteerd worden, maar ook geïntegreerd worden in het ERP-systeem als een module of dashboard.
* Het doel is dat klanten direct ondersteuning krijgen bij hun prognoses, bijvoorbeeld via waarschuwingen (“Deze prognose lijkt 80% kans te hoog”) of via dashboards die structurele afwijkingen tonen.
* Een ander doel is kennisoverdracht: zorgen dat interne teams en klanten begrijpen hoe de analyses tot stand komen en hoe zij de resultaten moeten interpreteren.
* Deployment draait dus niet alleen om implementatie, maar ook om borging, zodat de data-gedreven aanpak duurzaam voordeel oplevert.