Data Dictionary – Predictive Maintenance BI

Overview

This document describes the star schema data model used for the Predictive Maintenance Analytics project. The model consists of one fact table and three dimension tables.

Data Lineage

Source: AI4I 2020 Predictive Maintenance Dataset (UCI/Kaggle)
Raw File: `data/raw/predictive_maintenance.csv` (10,000 records)
Processed Files: `data/processed/` folder
ETL Process: `etl/etl_ai4i.ipynb`

Clean_Dataset (Fact_MachineEvents)

The file `clean_dataset.csv` contains the same fields as the `Fact_MachineEvents` table below.

- **EventID**: Surrogate key for each machine reading (integer).
- **DateID**: Foreign key to Dim_Date (YYYYMMDD integer).
- **MachineID**: Foreign key to Dim_Machine.
- **ProductTypeID**: Foreign key to Dim_ProductType.
- **AirTempC**: Air temperature around the machine in degrees Celsius.
- **ProcessTempC**: Process temperature during the operation in degrees Celsius.
- **RotationalSpeed**: Rotational speed of the tool in rpm.
- **Torque**: Torque applied during the operation in Nm.
- **ToolWear**: Cumulative tool usage in minutes.
- **MachineFailure**: 1 if any failure occurred during the cycle, 0 otherwise.
- **TWF**: Tool wear failure indicator (1/0).
- **HDF**: Heat dissipation failure indicator (1/0).
- **PWF**: Power failure indicator (1/0).
- **OSF**: Overstrain failure indicator (1/0).
- **RNF**: Random failure indicator (1/0).

Dim_ProductType

- **ProductTypeID**: Surrogate key for product type.
- **ProductType**: Product variant label (L, M, H).

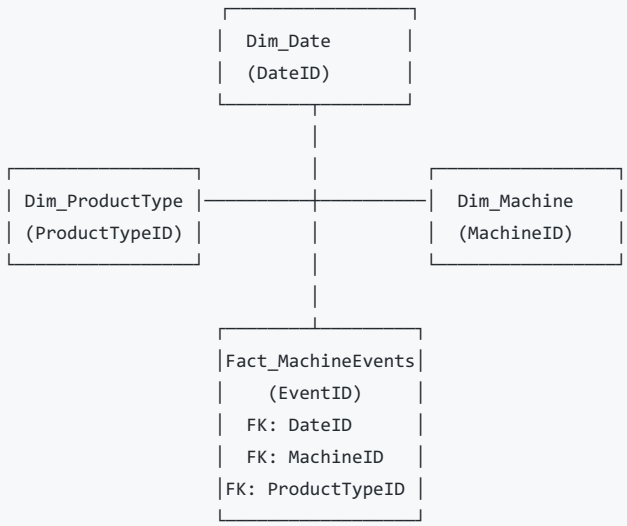
Dim_Date

- **DateID**: Surrogate key for the calendar date (YYYYMMDD).
- **Date**: Calendar date.
- **Day**: Day of month (1–31).
- **Month**: Month number (1–12).
- **Quarter**: Quarter of the year (1–4).
- **Year**: Year (e.g., 2025).
- **Shift**: Simple shift label (e.g., “Day”).

Dim_Machine

- **MachineID**: Surrogate key for the machine.
- **MachineName**: Human-readable machine name (Machine A, B, C).
- **Line**: Production line name.
- **Location**: Plant or site where the machine is installed.

Star Schema Relationships



Sample Values

Table	Field	Sample Values
Fact_MachineEvents	AirTempC	24.95, 25.05, 25.15
Fact_MachineEvents	RotationalSpeed	1408, 1551, 1667 rpm
Fact_MachineEvents	MachineFailure	0 (no failure), 1 (failure)
Dim_ProductType	ProductType	L (Low), M (Medium), H (High)
Dim_Machine	MachineName	Machine A, Machine B, Machine C
Dim_Date	DateID	20250101, 20250102, ...

Data Types Reference

Table	Field	Data Type	Nullable
Fact_MachineEvents	EventID	INTEGER	No (PK)
Fact_MachineEvents	DateID	INTEGER	No (FK)
Fact_MachineEvents	MachineID	INTEGER	No (FK)
Fact_MachineEvents	ProductTypeID	INTEGER	No (FK)
Fact_MachineEvents	AirTempC	DECIMAL(10,2)	No
Fact_MachineEvents	ProcessTempC	DECIMAL(10,2)	No
Fact_MachineEvents	RotationalSpeed	INTEGER	No
Fact_MachineEvents	Torque	DECIMAL(10,2)	No
Fact_MachineEvents	ToolWear	INTEGER	No
Fact_MachineEvents	MachineFailure	INTEGER (0/1)	No
Fact_MachineEvents	TWF, HDF, PWF, OSF, RNF	INTEGER (0/1)	No

Table	Field	Data Type	Nullable
Dim_Date	DateID	INTEGER	No (PK)
Dim_Date	Date	DATE	No
Dim_Date	Day, Month, Quarter, Year	INTEGER	No
Dim_Date	Shift	VARCHAR(20)	No
Dim_Machine	MachineID	INTEGER	No (PK)
Dim_Machine	MachineName	VARCHAR(50)	No
Dim_Machine	Line	VARCHAR(50)	No
Dim_Machine	Location	VARCHAR(50)	No
Dim_ProductType	ProductTypeID	INTEGER	No (PK)
Dim_ProductType	ProductType	VARCHAR(1)	No

ETL Transformations Applied

Transformation	Description
Column Renaming	`Air temperature` → `AirTempC`, `Machine failure` → `MachineFailure`
Temperature Conversion	Kelvin to Celsius: `AirTempC = AirTemperature - 273.15`
Surrogate Keys	Generated `EventID`, `DateID`, `MachineID`, `ProductTypeID`
Date Engineering	Created `Day`, `Month`, `Quarter`, `Year` from timestamps
Machine Assignment	Simulated 3 machines using `MachineID = (index % 3) + 1`
Data Type Casting	Failure indicators cast to INTEGER (0/1)

Business Glossary

Term	Definition
Machine Failure	Binary indicator (1/0) if any failure occurred during production cycle
TWF	Tool Wear Failure – failure due to excessive tool usage
HDF	Heat Dissipation Failure – failure due to temperature issues
PWF	Power Failure – failure related to power supply
OSF	Overstrain Failure – failure from exceeding load limits
RNF	Random Failure – unpredictable failure events
Product Type L/M/H	Low/Medium/High quality product variants
Tool Wear	Cumulative usage time of cutting tool (minutes)