Weather prediction



[**Introduction 3**](#_hvlomg2fq2tm)

[**What? 3**](#_6h3xu09qv4la)

[**Why? 3**](#_vcgu5am782gs)

[**Who? 3**](#_lxjv13y58r6z)

[**When? 3**](#_m72y2ei1zt26)

[**Domain understanding 4**](#_bku1sj94ec63)

[**Analytic approach 4**](#_rhcbr5yzqn2x)

[DOT-framework methods 4](#_3nj274xdn4z8)

[**Data requirements 4**](#_mgbkmk3i94h1)

# Introduction

## What?

This project focuses on developing a weather prediction model. Using historical and real-time meteorological data, we aim to forecast key weather parameters such as temperature, precipitation, humidity, and wind speed. The objective is to build a reliable model that can provide short-term weather predictions with high accuracy.

## Why?

Accurate weather forecasting is essential for various sectors, including transportation, agriculture, and event planning. Reliable predictions help in risk management, improving safety, and optimizing daily operations. Making precise forecasting crucial for residents and businesses. This project aims to improve existing weather prediction efforts using data-driven methods and machine learning techniques.

## Who?

The project involves "WeatherWise," a local weather service provider

## When?

The project will be executed in several phases:

1. **Phase 1 - Proposal (Tuesday)**: Creating the proposal, gathering historical weather data and identifying relevant datasets.
2. **Phase 2 - Provisioning (Wednesday):** Ensuring data quality by evaluating the imported data, performing exploratory analysis, and applying necessary preprocessing steps.
3. **Phase 3 - Predictions (Thursday)**: Training and testing machine learning models. Assessing model accuracy and making improvements.
4. **Presentation (Friday)**: Highlighting key challenges, solutions, and insights gained.

# Domain understanding

* Main research question: Can we develop an **accurate weather prediction** model for **Eindhoven** using **machine learning** techniques
* Methods of research: Data **collection** from meteorological sources (Open Weather API), data **analysis**, and model **evaluation**

# Analytic approach

* The target variable will be the weather conditions.
* Since “weather conditions” is a categorical target variable (e.g., sunny, rainy, cloudy), it falls under **classification**

## DOT-framework methods

Keep in mind that the DOT framework is being used as a guideline and not strictly being followed in this kick-off project due to time constraints:

* Field, Exploratory data analysis (ML): <https://ictresearchmethods.nl/field/exploratory-data-analysis/>
* Lab, Data analytics:

<https://ictresearchmethods.nl/lab/data-analytics/>

* Lab, Data quality check (ML):

<https://ictresearchmethods.nl/lab/data-quality-check/>

* Showroom, pitch:

<https://ictresearchmethods.nl/lab/data-quality-check/>

# Data requirements

Data points used:

**Temperature (°C/°F)** – Current air temperature

**Humidity (%)** – Relative humidity percentage

**Pressure (hPa)** – Atmospheric pressure

**Wind Speed (m/s or km/h)** – Speed of wind at different altitudes

**Precipitation (mm)** – Rainfall or snowfall amount per hour

Could be used for better results during later phases:

**Wind Direction (°)** – Direction of the wind

**Cloud Cover (%)** – Fraction of sky covered by clouds

**Dew Point (°C/°F)** – Temperature at which air becomes saturated with moisture

The data used was sourced from **Open-Meteo API** and it contains historical data from 2020-01-01 until the most recent date.

There are no any specific privacy or ethical concerns as the data is not related to a specific individual or groups of individuals and it is publicly available