Weather Forecast (Temperature) Prediction

Abstract:

Forecasting weather conditions is important for, e.g., operation of hydropower plants and for flood management. Mechanistic models are known to be computationally demanding. Hence, it is of interest to develop models that can predict weather conditions faster than traditional meteorological models. The field of machine learning has received much interest from the scientific community

Problem Statement:

Analyze the real historical weather and build a linear regression model with PySpark to predict the apparent temperature.

Variable Description:

The CSV file includes an hourly/daily summary for Szeged, Hungary area, between 2006 and 2016.

Column	Description
Formatted Date	Date of record
Summary	Weather summary
Precip Type	Categorical temperature whether there is rain or snow
Temperature	Air Temperature in celsius
Apparent Temperature (C)	The apparent temperature in celsius
Humidity	Humidity level
Wind Speed (km/h)	Wind speed in kilometre per hour
Wind Bearing (degrees)	"Wind Bearing" indicates the direction toward which wind is moving (Direction is in degrees ranging from 0 to 360 degrees)

Visibility (km)	The distance at which an object can be clearly discerned
Loud Cover	Cloud Cover
Pressure (millibars)	Atmospheric Pressure in millibars
Daily Summary	Daily weather report

Scope:

- Analyzing the existing data and getting valuable insights about the weather pattern
- Data pre-processing with PySpark
- Building a linear regression model in a distributed environment using Spark ML

Learning Outcome:

The students will get a better understanding of how the variables are linked to each other and will be able to apply linear regression with Spark ML that predicts the temperature forecast.