Tomorrow's Rain Forecast for New York

The Historical Hourly Weather Data 2012-2017 on Kaggle contains 5 years of hourly measurements of various weather attributes, such as temperature, humidity, air pressure, etc. The data are available to download from Kaggle at the following link:

https://www.kaggle.com/selfishgene/historical-hourly-weather-data

Task: Given historical weather data, can you predict whether *tomorrow* will be *rainy* in New York? Any weather description that is a form of liquid precipitation is considered rainy, such as those containing keywords "rain", "thunderstorm", "drizzle", etc. You have flexibility in defining "rainy" and "tomorrow" (e.g. next 24 hours from current hour); no matter what your definitions are, please make them clear in the report.

In the open-ended quiz, please make sure you address the following questions:

- Define metrics to evaluate algorithm performance
- What are your training, validation, and testing sets?
- Implement a simple baseline
- Compare the performance of one more algorithm to the baseline
- At least one algorithm should use a machine-learning/deep-learning approach

To simplify the problem, you can choose to use New York's data only, and ignore all other cities, however, you are not restricted to doing so.

Please submit your solution in either of the following formats:

Submit a Jupyter notebook detailing the analysis you performed. The notebook should include code, figures and/or tables summarizing results, and, importantly, rationale explaining why you chose any specific analytical methodology.

Alternatively, submit a Python script/package with all code written for model development and analysis. Include environment requirements/dependencies. Also, submit a description of each algorithm and discussion of performance. Include any figures/tables that are helpful to explain the result.

Guidelines:
- All models must be trained locally on a single workstation. Do not use computing clusters or cloud computing services.
- Please use data only contained in this dataset.
- This homework is open-ended, and is not supposed to take more than one day. No pressure to spend extra time trying to find the "best" solution.