ADS 506 Module 5 Exercises

Gabi Rivera

2023-11-23

1. A large medical clinic would like to forecast daily patient visits for purposes of staffing.

```
#a. If data are available only for the last month, how does this affect the
choice of model-based vs. data-driven methods?

# The limited available data maybe challenging for data-driven methods that's
more suited for substantial amount of data. So model-based might be the
better choice for a months worth of available data to forecast daily patient
visits. It's not a complex forecasting request so model-based will perform
well with this time series.

# b. The clinic has access to the admissions data of a nearby hospital. Under
what conditions will including the hospital information be potentially useful
for forecasting the clinic’s daily visits?

# Having access to hospital information will be highly useful for forecasting
the clinic's daily visits. So in this case the location is beneficial in a
way that hospital admission patterns may provide insights on the rate and
fluctuation of the clinics visiting volume. An increase in hospital admission
might take away from the clinics visit volume for transitioning patients or
patients with the same conditions out of the population. The hospitals
seasonal trends and holiday cycles can also possibly influence the clinic's
visit volume.

# c. Thus far, the clinic administrator takes a heuristic approach, using the
visit numbers from the same day of the previous week as a forecast. What is
the advantage of this approach? What is the disadvantage?

# Heuristic approach is advantageous in a clinical setting for its
simplicity. It can be done quickly with less use of data intensive analysis.
The forecast is a straight forward result that can capture weekly patterns.
The disadvantage is that this approach is not robust and will not be able to
adapt or be sensitive enough to capture unexpected variability in visits. It
will not be as accurate as a model over time in comparison.

# d. What level of automation appears to be required for this task? Explain.

# A moderate automation might be a better choice in this scenario. That's
because the need to forecast daily visit is not as complex. A basic automated
time series model can be implemented to have an improvement against the
```

```
heuristic approach without being overtly complicated. And at the same time,
the need for a more accurate forecasting is satisfied in order to supply
staffing needs of the clinic.

#e. Describe two approaches for improving the current heuristic (naive)
forecasting approach using ensembles.

# One ensemble approach is to combine the naive forecasting of the clinical
data with the nearby hospital admission data. Another approach can be to
combine individual naive forecast using simple averaging aggregation approach
```

2. The ability to scale up renewable energy, and in particular wind power and speed, is dependent on the ability to forecast its short-term availability. Soman et al. (2010) describe different methods for wind power forecasting (the quote is slightly edited for brevity):

Persistence Method: This method is also known as "Naive Predictor". It is assumed that the wind speed at time t + δt will be the same as it was at time t. Unbelievably, it is more accurate than most of the physical and statistical methods for very-short to short term forecasts.

Physical Approach: Physical systems use parameterizations based on a detailed physical description of the atmosphere.

Statistical Approach: The statistical approach is based on training with measurement data and uses difference between the predicted and the actual wind speeds in the immediate past to tune model parameters. It is easy to model, inexpensive, and provides timely predictions. It is not based on any predefined mathematical model and rather it is based on patterns.

Hybrid Approach: In general, the combination of different approaches such as mixing physical and statistical approaches or combining short term and medium term models, etc., is referred to as the hybrid approach.

```
# a. For each of the four types of methods, describe whether it is model-
based, data-driven, or a combination .

# Persistence method: data driven because it assumes the wind speed will be
the same as the current wind speed.
# Physical approach: model based because it relies on physical description of
the atmosphere.
# Statistical approach: data driven because it captures patterns from data
and measurements.
# hybrid approach: combination because it combines different methods to
improve on forecasting.

# b. For each of the four types of methods, describe whether it is based on
extrapolation, causal modeling, correlation modeling, or a combination.
```

```
# Persistence method: exploration because it assumes a stationary can be
extended to the future and ignores other parameters that may influence the
outcome.
# Physical approach: casual or correlation modeling because it relies on
physical aspects to understand whether or not the relationship with the wind
pattern is casual or due to correlation.
# Statistical approach: extrapolation modeling because it uses historical
information to predict wind speed.
# hybrid approach: combination because it combines methods to leading to an
ensemble.

# c. Describe the advantages and disadvantages of the hybrid approach.

# For advantages, combining forecasting models improves on accuracy and
robustness. A high performing model can contribute to pick up the slack of a
poor performing model under certain conditions. This contributes to hybrid
model's adaptability and reduction of model result uncertainties.
# For disadvantage, hybrid approach can be too complex and uses a lot of
computational resources. There's also a challenge in integrating models
seamlessly and more data might be required for processing along with
challenges in interpreting all the components that goes into forecasting.
```