Applied Stochastic Assignment 5

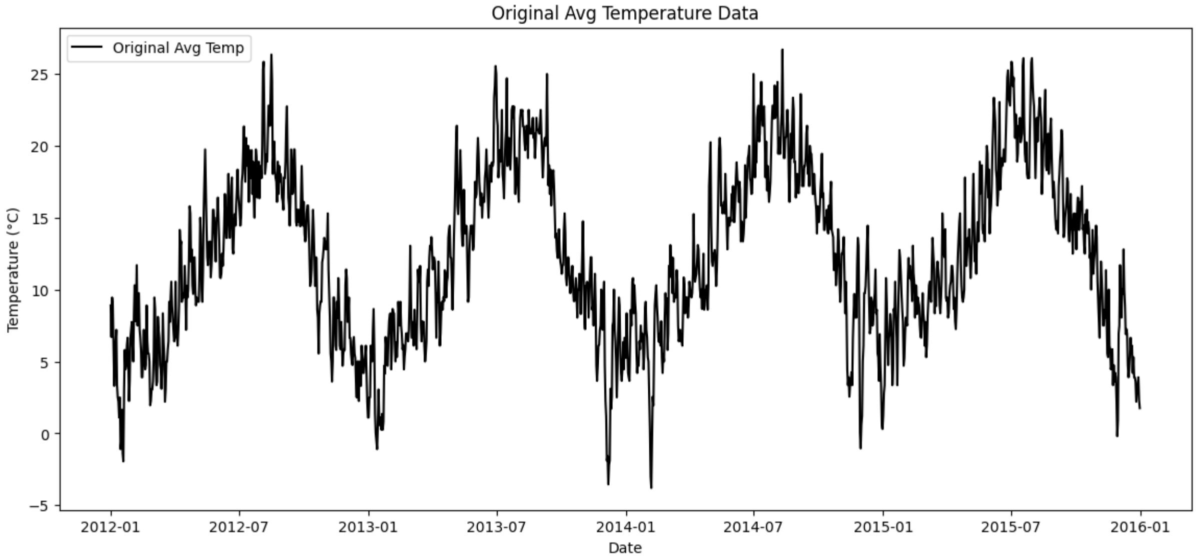
Manyara Baraka - mbaraka

November 28, 2024

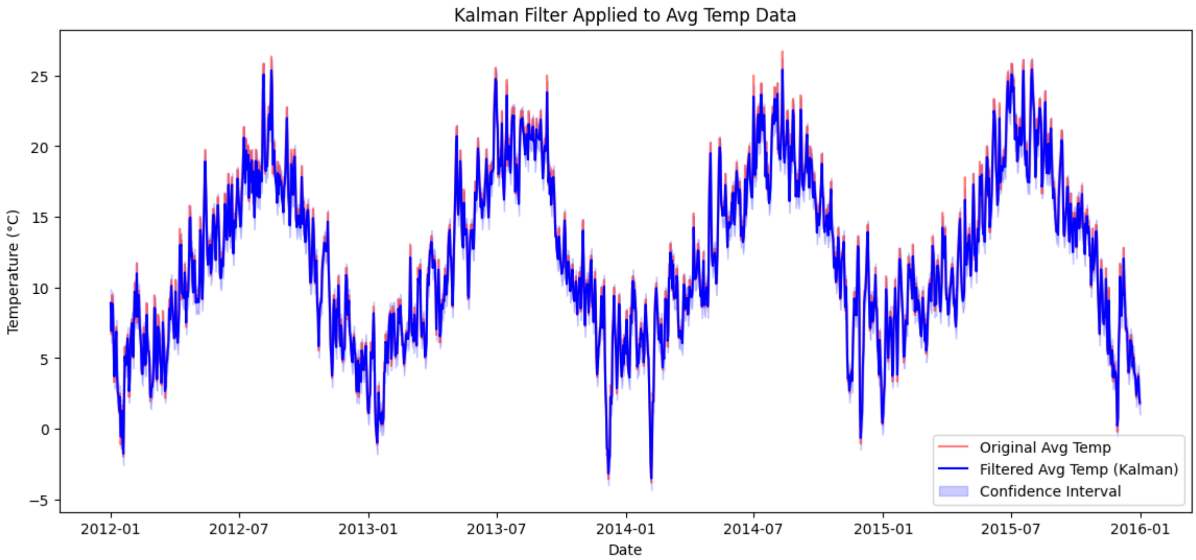
1 Kalman Filters and Hidden Markov Models (HMMs) in Time Series Analysis

[A]. Kalman Filter

The original Temperature graph



Kalman Filtered temperature



Perfomance Metrics

Performance Metric	Value
Mean Squared Error (MSE)	0.1111
Root Mean Squared Error (RMSE)	0.3334
Mean Absolute Error (MAE)	0.2635

Note: the above observation was made using the following values of parameter:

- $\alpha = 1$
- $q = 2$
- $r = 1$

This means that the system will rely more on the measurements and less on the model prediction.

How well did the Kalman Filter smooth the temperature data and predict future temperatures?

The low values of the performance metric indicate the high accuracy of Kalman in smoothing the data. Additionally, it reduces noise which allows the Kalman filter to understand the data's trends clearly.

From the second graph Kalman filter shows a potential capability for smoothing the temperature data effectively removing noise as it preserves the underlying trends. The filtered line blue follows the original data without overfitting.

Was the confidence interval effective in depicting the variability?

Yes, this is shown by:

- The light blue-shaded region in the second graph (filters graph), accurately represents uncertainty around the smoothed temperature estimates.
- The intervals align closely with the fluctuations in the original data, this indicates the confidence interval indicating the Kalman robustness in handling uncertainty.

[B]. HMM and Viterbi Algorithm

Training Set Comparison

	Actual	Predicted
0	drizzle	sun
1	rain	sun
2	rain	sun
3	rain	sun
4	rain	rain

Testing Set Comparison

	Actual	Predicted
0	rain	sun
1	rain	sun
2	rain	sun
3	rain	sun
4	sun	sun

HMM Performance

- **Training Set Accuracy:** 20.03%
- **Testing Set Accuracy:** 19.11%

How accurately did the HMM predict the hidden weather states?

From the accuracy values the HMM struggles to accurately predict the hidden weather states.

What were the limitations of using an HMM for this type of data?

- **Simple state representation**

  Some states like rain or drizzle tend to overlap and are influenced by other factors like temperature or humidity. The hidden Markov model struggled to differentiate the states with fewer features in the dataset.

- **Lack of features**

  The data didn't have additional features such as wind speed, humidity, and others, that could inform the model better. The features might guide the HMM to also identify distinct weather states more accurately.

- **Poor generalization**

  The model had poor generalization in both the testing and the training datasets.

[C]. Comparison

Which model provided better predictive performance for the weather data?

Kalman Filter: The Kalman filter smoothed the continuous temperature data perfectly and it provided a reliable confidence interval.

Under what circumstances might each method be preferable?

(a) **Kalman Filter**

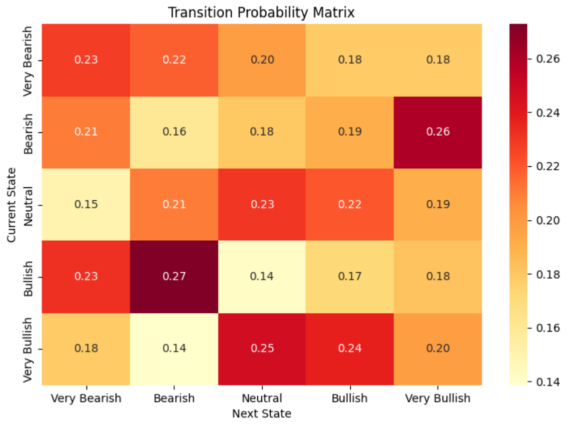
- Continous time-series with noise such as stock price and temperatures
- Systems that need real-time updates and forecasting
- where the confidence intervals and uncertainty are critical.

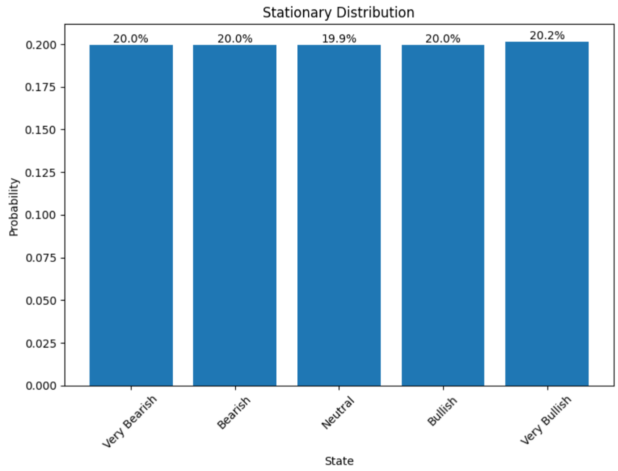
(b) **HMM**

- Discrete classification problems where the states have a clear transition such as as speech recognition or disease progression.
- Systems with more features and well-defined state-to-state transition structure.

2 Markov Chain for Stock Price Prediction

Transition Matrix





	Very Bearish	Bearish	Neutral	Bullish	Very Bullish
Very Bearish	0.227723	0.217822	0.198020	0.178218	0.178218
Bearish	0.210000	0.160000	0.180000	0.190000	0.260000
Neutral	0.150000	0.210000	0.230000	0.220000	0.190000
Bullish	0.232323	0.272727	0.141414	0.171717	0.181818
Very Bullish	0.178218	0.138614	0.247525	0.237624	0.198020

The matrix indicates how frequently the stock moved from one state to another over time.

For the matrix:

- **Rows** Represent the current state
- **Columns** Represent the next state
- **Values** represent the probability of the transition from the current state to the next state.

The high diagonal probabilities show that the stocks remain in the same state over time. For instance the for **Very bearish to bearish: 22.8%** whereas the low diagonal probabilities like **Bullish to Bullish: 17.2%** this indicates a short period where the market will persist over a short period.

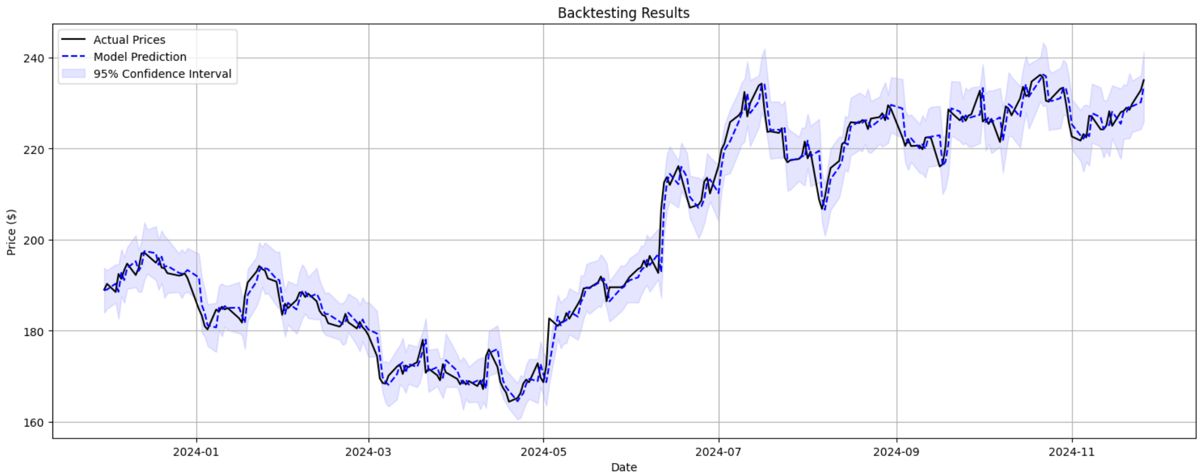
Stationary Distribution

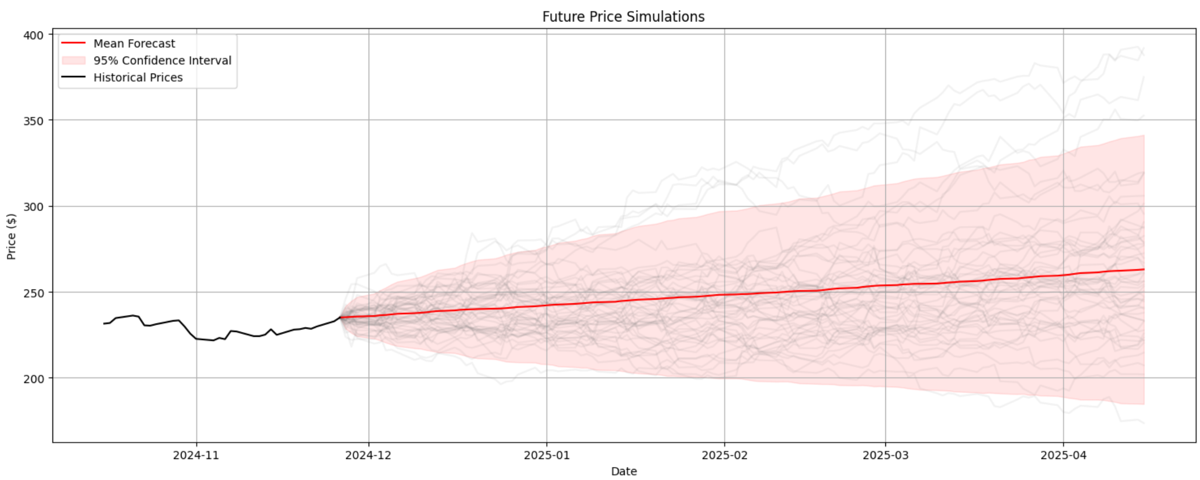
State	Probability
Very Bearish	0.199616
Bearish	0.199703
Neutral	0.199484
Bullish	0.199585
Very Bullish	0.201611

The stationary distribution represented the long-term probability of the stock being at different and each state.

The distribution had an almost uniform distribution for each of the states, this indicates that the stock has no inherent bias towards a given state in a long time. Additionally, it indicates that overtime in each state is likely to happen over a period of time.

Backtesting and Future Simulation





Backtesting Results

After 2024-07, the model seems to have strong accuracy in tracking the price. Additionally, the 95% confidence interval effectively captures the most price movements.

Future Price simulation

The mean forecast indicates a gradual upward trend from the current price of $ 235.06.The confidence interval, the "pink" area, widens significantly. This indicates the model's robustness as it covers most of the historical prices. The widening of the confidence also tends to explain the increasing uncertainty of the prices over a long time.

Future Prediction Statistics

Metric	**Value**
Current Price	$235.06
Predicted Price (in 100 days)	$263.05
Average Predicted Price (100 days)	$263.05
Prediction Range	$145.59 - $425.79
Historical Volatility (%)	1.41%
Simulated Volatility (%)	10.95%
90% Confidence Interval (Final Price)	$203.20 - $329.48
95% Confidence Interval (Final Price)	$196.56 - $347.90

Model evaluation

The Markov chain model generally demonstrates a strong predictive capability as it highlights key risk considerations such as:

- The wide prediction range shows uncertainty in the long-term forecasts.
- The model shows a market without strong directional distribution bias through the balances stationary distribution.
- It also captures effectively both the trend persistence and the mean reversion tendencies.