**"Time Series Analysis and Forecasting of Snow Depth in Tupper Lake Using ARIMA,** **AutoARIMA and HoltWinters"**

Contents

[Introduction 2](#_Toc193079039)

[Data Preprocessing 2](#_Toc193079040)

[Visualization: 3](#_Toc193079041)

[Model Building 5](#_Toc193079042)

[Model Evaluation: 7](#_Toc193079043)

[Conclusion: 7](#_Toc193079044)

**Rahul Tammalla**

**Clarkson University**

# Introduction

**Objective:**

The primary objective of this project was to forecast the snow depth in “Tupper Lake, NY”, using time series forecasting models. These models were used to predict snow depth for the upcoming 12 months, helping to better understand seasonal patterns and prepare for future winter conditions.

**Importance of the Problem:**

Snow depth forecasting plays a critical role in managing winter conditions in regions like Tupper Lake. Accurate snow depth predictions are valuable for several reasons:

* **Infrastructure Management**: Planning for snow removal and road safety.
* **Winter Sports**: Accurate forecasting aids in decision-making for winter tourism.
* **Disaster Preparedness**: Helps in predicting snow-related hazards such as avalanches or flooding.

**Models Used:**

* **ARIMA** (AutoRegressive Integrated Moving Average): A traditional time series model.
* **AutoARIMA**: An automated version of ARIMA that selects the best parameters.
* **HoltWinters Exponential Smoothing**: A method for capturing both trend and seasonality.

# Data Preprocessing

**Data Description:**

The dataset used in this project contains hourly snow depth measurements, along with other meteorological features like temperature, humidity, solar insolation, and precipitation. The snow depth is the target variable for forecasting.

Head of Data:

A screenshot of a computer

AI-generated content may be incorrect.

* **Handling Missing Data:**

Missing values in the snow\_depth column were handled by forward filling, which propagated the last valid measurement to replace any missing snow depth entries. Additionally, non-positive snow depth values (<= 0) were set to 0 for consistency in the dataset.

* **Time Formatting:**

The time column was converted to the datetime format to enable proper time series analysis and ensure that time-based indexing worked correctly for the models.

# Visualization:

**Snow depth over the years (2017- 2025)**

A graph with blue lines

AI-generated content may be incorrect.

A graph of different colored lines

AI-generated content may be incorrect.

**Snow depth over the course of each year** from **2022 to 2025**

A graph of a graph of a graph

AI-generated content may be incorrect.

**Yearly Observations**:

* **2022**: Snow depth is **high during the beginning of the year** (around days 0-100), gradually decreases, and then sees another peak toward the end of the year, likely corresponding to another snow accumulation period.
* **2023**: Similar pattern to 2022, but with more **frequent variations** and **larger fluctuations** in snow depth throughout the year.
* **2024**: Exhibits a similar pattern, but the snow depth does not show as much variation during the middle of the year, likely indicating a milder winter or lesser snowfall.
* **2025**: **Rapid increase** in snow depth in the early part of the year, showing a significantly higher snow accumulation as compared to the previous years.

**Interpretation**:

* This visualizations is helpful for understanding **year-over-year** variations in snow depth.
* The plots shows that **snow depth tends to rise in the winter** and **decline in the summer**, which is expected due to the seasonal nature of snow accumulation and melting.
* The **early months of each year** typically show the most **significant snow accumulation** as winter progresses.

# Model Building

**ARIMA Model:**

The ARIMA model was applied with parameters (p = 2, d = 1, q = 1) and seasonal components (P = 1, D = 1, Q = 1) with a seasonal period of 12 months. This model is suitable for datasets with both trend and seasonal components, capturing patterns in the snow depth data.

**AutoARIMA Model:**

The AutoARIMA model was used to automatically select the optimal ARIMA parameters by testing various combinations of p, d, and q. This approach helps find the best model without manual parameter tuning.

**HoltWinters Model:**

The HoltWinters Exponential Smoothing model was applied to account for both trend and seasonality in the snow depth data. HoltWinters is effective in modeling time series with clear seasonal patterns and trends, making it a strong candidate for forecasting snow depth.

**Code:**

A screenshot of a computer program

AI-generated content may be incorrect.

**Snow Depth Forecasting using ARIMA, AutoARIMA, and HoltWinters**

A graph showing a graph

AI-generated content may be incorrect.

**Interpretation:**

* **Actual Snow Depth (Blue Line)**: Represents the recorded snow depth from 2017 to 2025, showing clear seasonal patterns with peaks in winter and valleys in warmer months.
* **ARIMA Forecast (Green Line)**: ARIMA captures the trend but doesn't handle sharp fluctuations well, especially for 2025, where it smooths out the peaks and valleys.
* **AutoARIMA Forecast (Yellow Dashed Line)**: AutoARIMA optimizes ARIMA's parameters automatically, providing a better fit than ARIMA but still struggling with sharp peaks, particularly in 2025.
* **HoltWinters Forecast (Red Dashed Line)**: HoltWinters does well in capturing the seasonal trend but fails to capture abrupt increases or decreases, like in 2025.

**Key Takeaway:**

* **ARIMA** and **AutoARIMA** provide reasonable trend forecasts but fail to capture sudden snow depth changes.
* **HoltWinters** is good at capturing seasonality but underperforms during rapid snow depth changes, especially in 2025.

# Model Evaluation:

**Evaluation Metrics:**

The models were evaluated based on two key metrics:

1. **MAE (Mean Absolute Error)**: Measures the average magnitude of the errors between predicted and actual values.
2. **RMSE (Root Mean Squared Error)**: Measures the square root of the average squared differences between predicted and actual values.

**Model Performance:**

* **ARIMA**:
  + **MAE**: 7.71
  + **RMSE**: 8.50
* **AutoARIMA**:
  + **MAE**: 3.39
  + **RMSE**: 5.21
* **HoltWinters**:
  + **MAE**: 2.65
  + **RMSE**: 3.95

**Interpretation**:

* **HoltWinters** outperformed both **ARIMA** and **AutoARIMA** in terms of accuracy, as indicated by the lowest **MAE** and **RMSE** values.
* **AutoARIMA** was more accurate than **ARIMA** but still did not match the performance of **HoltWinters**.

# Conclusion:

In this project, we applied three time series forecasting models **ARIMA**, **AutoARIMA**, and **HoltWinters** to predict snow depth in Tupper Lake, NY. The models were evaluated for their ability to capture seasonal patterns and fluctuations in snow depth.

* **ARIMA** and **AutoARIMA** provided reasonable forecasts but struggled with sharp fluctuations in snow depth, especially in years with rapid changes like 2025.
* **HoltWinters** captured the seasonality well but failed to predict sudden increases or decreases in snow depth.

Overall, the models offered valuable insights into snow depth trends, with **AutoARIMA** providing the most optimized results. However, further refinement or combining models may be needed to improve accuracy, particularly for sudden snow depth changes.