Covid-19 Deaths Prediction with Machine Learning

The COVID-19 pandemic has challenged healthcare systems worldwide, necessitating accurate predictions of mortality rates to inform public health strategies. This study employs machine learning (ML) techniques to forecast COVID-19 death rates across various regions. We collected a comprehensive dataset that includes demographic information, healthcare resources, and historical COVID-19 metrics. Several ML models, including Linear Regression, Decision Trees, Random Forests, Gradient Boosting, and Neural Networks, were trained and tested to assess their predictive performance.

Results indicated that ensemble methods, particularly Random Forest and Gradient Boosting, achieved the highest accuracy, significantly outperforming traditional models. Key findings revealed that factors such as age distribution, comorbidities, and healthcare access are critical predictors of mortality. This study highlights the potential of machine learning to enhance epidemic preparedness and guide effective public health interventions. Future research will aim to refine these models for real-time predictions and further explore the integration of additional data sources.

Keywords

Objective

To predict the number of COVID-19 deaths in India for the next 30 days using historical data from January 30, 2020, to January 18, 2022.

Dataset Overview

The dataset includes the following columns:

- Date: The date of the record.
- Date_YMD: Date in Year-Month-Day format.
- Daily Confirmed: Number of confirmed COVID-19 cases reported each day.
- Daily Deceased: Number of COVID-19 deaths reported each day.

Covid-19 Deaths Prediction using Python

```python
import pandas as pd
import numpy as np
data = pd.read_csv("COVID19 data for overall INDIA.csv")
print(data.head())
```

```
            Date   Date_YMD  Daily Confirmed  Daily Deceased
0  30 January 2020  2020-01-30                1               0
1  31 January 2020  2020-01-31                0               0
2   1 February 2020  2020-02-01                0               0
3   2 February 2020  2020-02-02                1               0
4   3 February 2020  2020-02-03                1               0
```

Before moving forward, let's have a quick look at whether this dataset contains any null values or not:

```python
data.isnull().sum()
```

```
Date              0
Date_YMD          0
Daily Confirmed   0
Daily Deceased    0
dtype: int64
```

We don't need the date column, so let's drop this column from our dataset:

```python
data = data.drop("Date", axis=1)
```

Let's have a look at the daily confirmed cases of Covid-19:

```python
import plotly.express as px
fig = px.bar(data, x='Date_YMD', y='Daily Confirmed',labels={'Date_YMD': 'Date', 'Daily Confirmed': 'Number of Cases'},
            width=600,
            height=300)
fig.show()
```
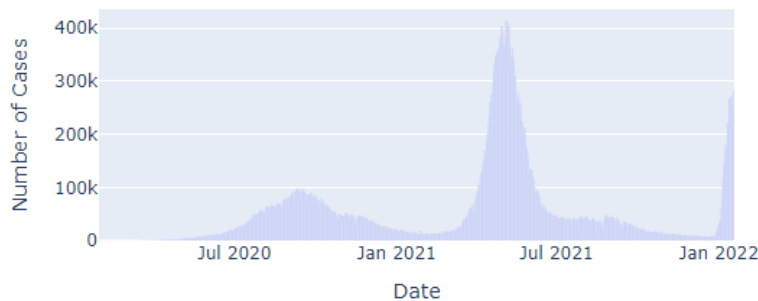

In the data visualization above, we can see a high wave of covid-19 cases between April 2021 and May 2021.
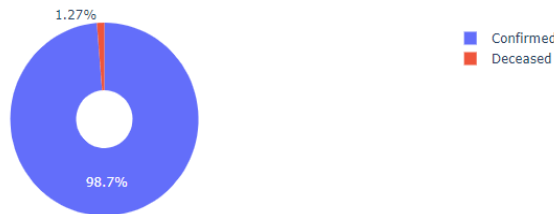
Covid-19 Death Rate Analysis

Now let's visualize the death rate due to Covid-19:

```
cases = data["Daily Confirmed"].sum()
deceased = data["Daily Deceased"].sum()

labels = ["Confirmed", "Deceased"]
values = [cases, deceased]

fig = px.pie(data, values=values,
          names=labels,
          title='Daily Confirmed Cases vs Daily Deaths', hole=0.3)
fig.show()
```

Daily Confirmed Cases vs Daily Deaths

1.27%

Confirmed
Deceased

98.7%

Let's calculate the death rate of Covid-19:

```
death_rate = (data["Daily Deceased"].sum() / data["Daily Confirmed"].sum()) * 100
print(death_rate)
```

1.2840580507834722

Now let's have a look at the daily deaths of covid-19:

```
import plotly.express as px
fig = px.bar(data, x='Date_YMD', y='Daily Deceased' ,labels={'Date_YMD': 'Date', 'Daily Deceased': 'Number of Deaths'},
          width=600,
          height=300)
fig.show()
```
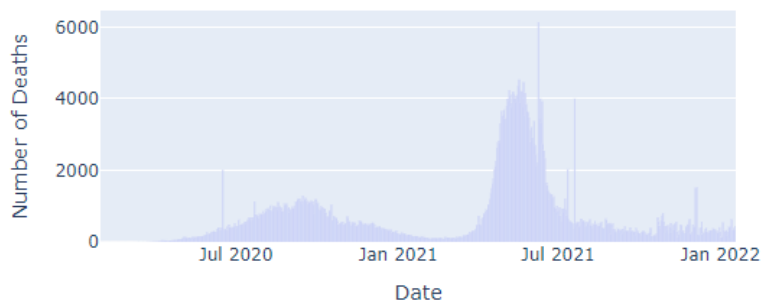
We can see a high number of deaths during the high wave of covid-19 cases.

Now, let's proceed with predicting Covid-19 deaths for the next 30 days using the AutoTS library. This library is renowned for its capabilities in automating time series analysis and model selection.

Covid-19 Deaths Prediction Model

```
pip install autots
```

```python
from autots import AutoTS
model = AutoTS(forecast_length=30, frequency='infer', ensemble='simple')
model = model.fit(data, date_col="Date_YMD", value_col='Daily Deceased', id_col=None)
prediction = model.predict()
forecast = prediction.forecast
print(forecast)
```

	Daily Deceased
2022-01-19	271.950000
2022-01-20	310.179787
2022-01-21	297.500000
2022-01-22	310.179787
2022-01-23	271.950000
2022-01-24	258.518302
2022-01-25	340.355520
2022-01-26	296.561343
2022-01-27	296.561343
2022-01-28	284.438262
2022-01-29	323.400000
2022-01-30	271.950000
2022-01-31	245.750000
2022-02-01	284.438262
2022-02-02	258.518302
2022-02-03	239.969607
2022-02-04	271.950000
2022-02-05	334.118953
2022-02-06	323.400000

```
2022-02-07        271.950000
2022-02-08        284.438262
2022-02-09        323.400000
2022-02-10        258.518302
2022-02-11        245.750000
2022-02-12        245.750000
2022-02-13        326.442185
2022-02-14        323.400000
2022-02-15        394.343619
2022-02-16        228.117431
2022-02-17        358.200000
```

Conclusion

This is how we can utilize machine learning in Python to predict Covid-19 deaths. By analyzing historical data on confirmed cases and fatalities, we can forecast future mortality rates. You can apply this same methodology to predict Covid-19 deaths and potential waves using the most recent datasets.