Armed Conflict In The USA

Jefferey Lin

The Issues:

In the United States, despite its status as an advanced economy and a first-world nation, social unity often seems elusive, particularly evident through the deep political division. As 2024 in an election year, these divisions are poised to deepen further. Beyond political differences, societal conflicts often stem from issues of racial and ethnic identity, which exacerbates tension.

The contentious debate between the left and right surrounding the First and Second Amendments fuel the fire. The right to freedom of speech and press, along with the right to bear arms, forms a battleground of ideological clashes. Some advocate for stricter gun policies in response to the mass shootings and urban violence, while others perceive the right to bear arms as integral to individual safety and autonomy.

Social shifts, such as the growing support for LGBTQ+ rights, highlight the stark contrast in values across the political spectrum. Growing environmental concerns, particularly regarding climate change, have galvanized activities to push for better conservation and sustainable practices. Events such as the overturning of Roe vs Wade led to further demonstrations concerning female empowerment and reproductive rights. Further socioeconomic disparities also contribute to significant social unrest. Economic inequality, lack of access to healthcare, housing insecurity and education inequality all drive marginalized communities to protest for equitable opportunities.

These multifaceted tensions have led to a surge in demonstrations across the country, as people mobilize to voice their grievances and advocate for change. The complex interplay of political, social and economic factors underscores the challenges facing the US with the goal of unity amidst diversity.

The Findings:

In the period from 2020 to 2023, the number of deaths resulting from violent demonstrations exhibited a steady decline, with 69 fatalities in 2020, decreasing to 45 in 2021, 41 in 2022, and 36 in 2023. Concurrently, the frequency of protests saw a significant reduction, with 1045 recorded in 2020, dropping to 307 in 2021, 247 in 2022, and 217 in 2023. Analyzing the plotted demonstration points on a map revealed that major cities like Los Angeles, Washington DC, and New York City had the highest density of demonstrations.

Notably, May 31st, 2020, marked the day with the highest number of demonstrations in a single day, totaling 103 demonstrations. Interestingly, the top five days, and seven of the top 10 days, with the highest number of demonstrations all occurred consecutively, surrounding the George Floyd murder. The other three days with the higher number of demonstrations all occurred in 2020.

Despite 2020 recording the highest number of demonstrations, surpassing the total of the other years combined, it had the lowest fatality-to-protest ratio among the four years. In 2020, this ratio was only 6.6%, significantly lower than the ratios observed in 2021, 2022, and 2023, which more than doubled it.

After developing a predictive model using the dataset, the results indicate that by analyzing the three nearest data points to predict whether a 2023 demonstration would lead to a fatality, we achieved an accuracy of 83%. This suggests that our model accurately predicted the presence or absence of fatalities in 2023 demonstrations 83% of the time.

Discussion:

On May 25, 2020, a highly politicized incident sparked widespread criticism of the police force following the death of George Floyd. Floyd was arrested outside a grocery store and had his neck pressed on by the knee of a police officer, Derek Chauvin, for nine and a half minutes. His death was ruled a homicide, leading to global outrage, particularly due to Chauvin being white. May 28th to 31st, as well as July 1st, 2nd, and 6th were among the top 10 days in terms of the number of demonstrations recorded in this dataset, coinciding with the aftermath of the George Floyd incident. Moreover, the deaths of black citizens such as Ahmaud Arbery and Breonna Taylor prompted widespread protests, with the hashtag #blackouttuesday gaining momentum on social media platforms worldwide.

The year 2020 also marked the onset of the pandemic, leading to increased isolation and heightened scrutiny of government policies. On July 25th, one of the top 10 days with the most demonstrations, the CDC's decision to reopen schools shocked parents nationwide.

In addition, the Senate acquitted President Donald Trump of impeachment charges in 2020, fueling further unrest, particularly among left-leaning individuals. These significant events contributed to the higher number of demonstrations in 2020 compared to all the remaining years combined.

Appendix A: Method

This project utilized a dataset sourced from the Armed Conflict Location & Event Data Project (ACLED), which consists of records of violent demonstrations in the US spanning from the 1st of January, 2020 to the 1st of January, 2024. In this data set, there were 191 demonstrations that resulted in fatalities and 1625 demonstrations with no fatalities recorded. The objective of this study is to predict whether a violent demonstration in 2023 resulted in a fatality, using training data from 2020 to 2022.

For background context, several maps were created to visualize the geographical distribution of these demonstrations and their fatality outcomes. Additionally, a time series chart was generated to illustrate the temporal trends in the number of demonstrations over the study period. Furthermore, a bar chart was constructed to analyze the frequency of protests each year, providing insight into the factors driving protest activity.

K-nearest neighbors (KNN) algorithm was employed for training and testing the data. To streamline the dataset, columns representing the year, month, and day were consolidated into a single column and the original columns were subsequently dropped. Longitude and latitude values were standardized to ensure consistency in distance calculations.The dataset was then divided into training and testing subsets. The training data encompassed demonstrations from 2020, 2021, and 2022, while the testing data, used for predicting fatalities in 2023 events, comprised demonstrations from that year. Selecting an appropriate value for K is crucial, as it determines the number of nearest neighbors considered for predictions. Odd values for K are preferred to avoid ties in voting. To account for the curvature of the Earth, the Haversine distance metric was incorporated, which considers both distance and curvature in distance calculations.

Finally, the KNN model predicted whether a fatality occurred in the 2023 events and provided accuracy metrics, including precision, recall, and F1-score. To visualize the impact of different K values, a plot depicting K against various metrics was constructed. Overall, this methodological approach aimed to provide insights into the factors influencing violent demonstrations and to develop a predictive model for fatalities based on historical data.

Appendix B: Results

The following three graphs, Figures 1, 2, and 3, plot the longitude and latitude points provided in the dataset to visualize the locations of all 1816 demonstrations that occurred in the US. Figure 1 displays all the demonstrations, while Figure 2 specifically plots demonstrations without fatalities. Conversely, Figure 3 illustrates demonstrations that resulted in fatalities.

Figure 1: Total Demonstrations

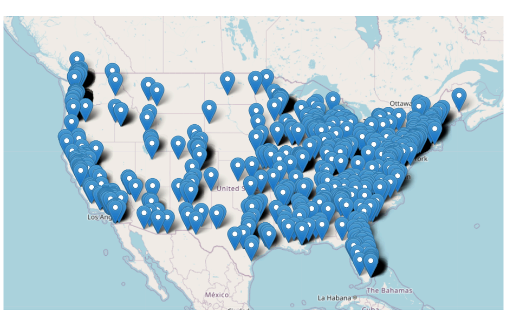


Figure2: Without Fatalities

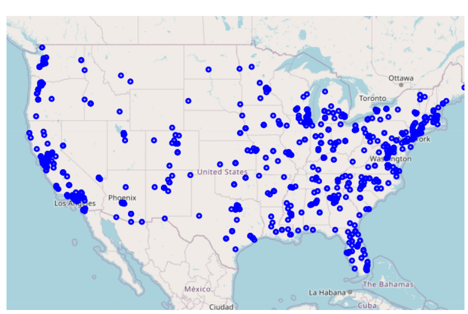
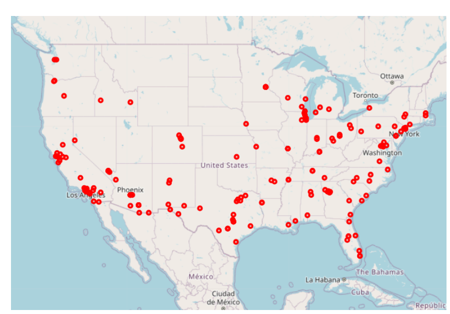


Figure 3: Fatalities



Figures 4 and 5 illustrate that 2022 stood out as the most demonstrated and politically polarizing year, also witnessing the highest number of fatalities. However, despite the high number of demonstrations, the fatality-to-protest ratio was approximately 6.6% for 2020, significantly lower than the 14.5% observed in 2021, and the 16.6% recorded for both 2022 and 2023.

Figure 4: Number of deaths per year.

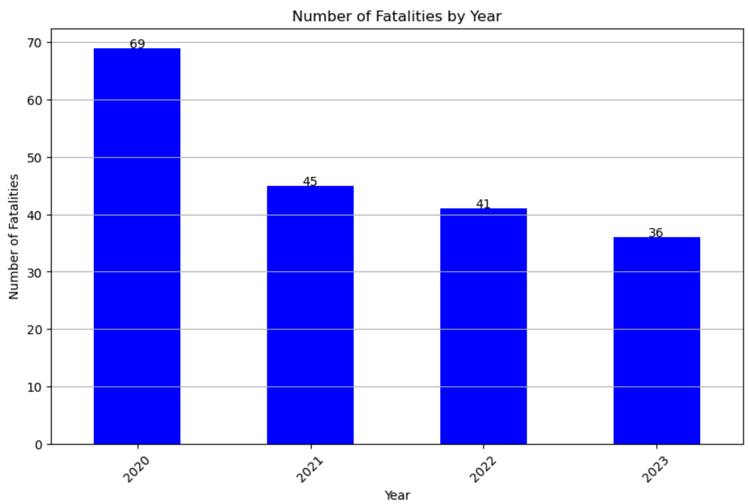


Figure 5: Number of protests each year

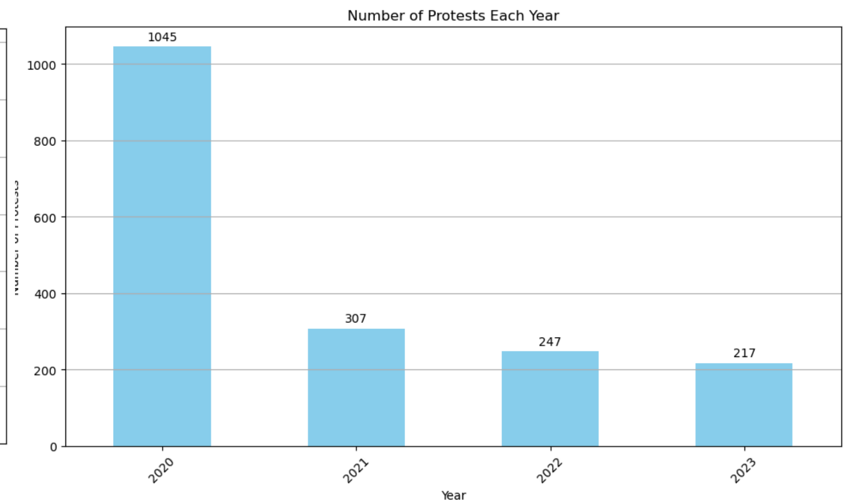
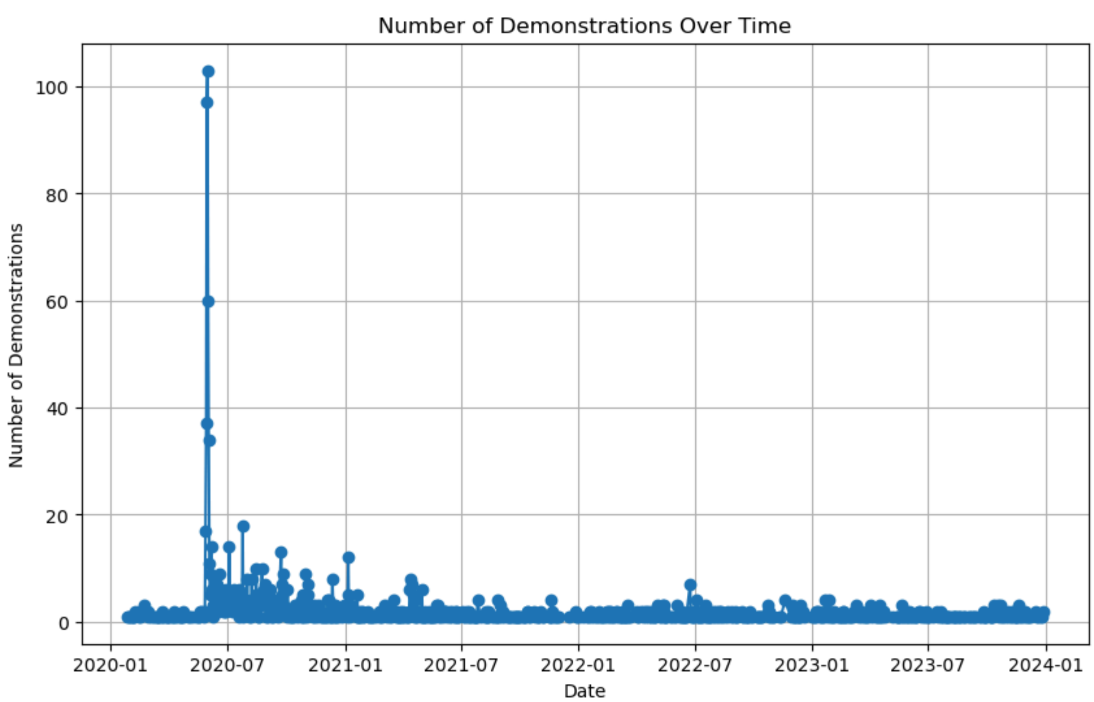


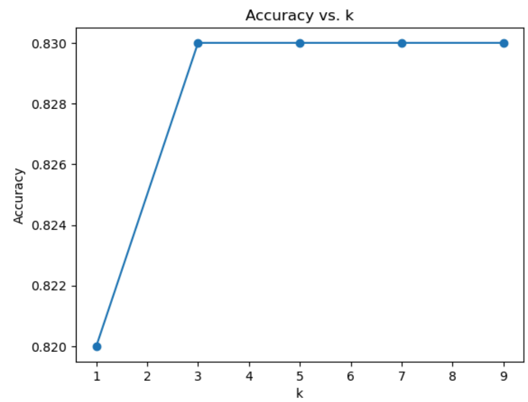
Figure 6 highlights significant events that occurred in 2020, leading to a notable spike in the number of demonstrations, including the COVID-19 pandemic and the death of George Floyd. In contrast, the plot for the other years remained relatively flat.

Figure 6: Demonstrations from 2020 to 2023



The predictive model's performance was evaluated using different values of k in the K-nearest neighbors algorithm. For k=1, the model achieved an accuracy of 82%, indicating that it correctly predicted the outcome of demonstrations 82% of the time. As the value of k increased to 3, 5, 7, and 9, the accuracy remained consistently high at 83%. This suggests that the model's ability to classify demonstrations as having a fatality or not was robust across different values of k, with an accuracy rate of 83% for the majority of tested scenarios.

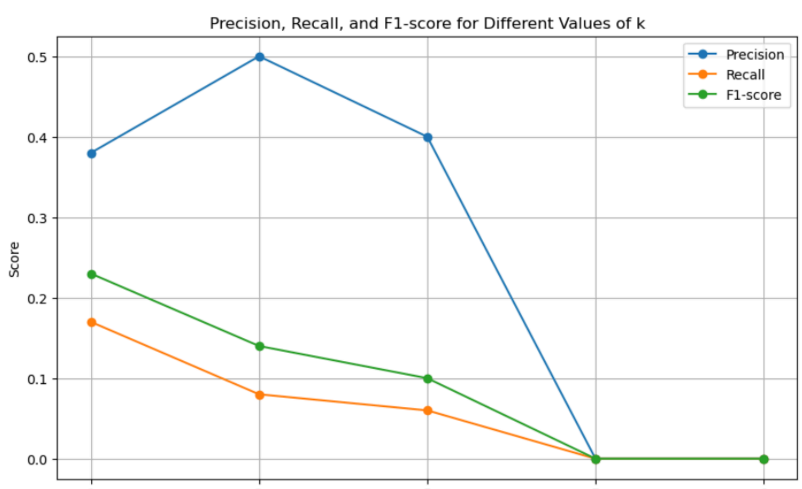
```
Accuracy for k=1: 0.82
Accuracy for k=3: 0.83
Accuracy for k=5: 0.83
Accuracy for k=7: 0.83
Accuracy for k=9: 0.83
```



The performance of the predictive model was evaluated using various values of k in the K-nearest neighbors algorithm. For k=1, the precision, recall, and F1-score were 0.38, 0.17, and 0.23, respectively. As the value of k increased to 3 and 5, precision and recall exhibited slight fluctuations, with precision peaking at 0.50 for k=3 and gradually decreasing for higher values of k. However, recall showed a decreasing trend across all values of k. Consequently, the F1-score, which balances precision and recall, also decreased as the value of k increased, indicating a trade-off between precision and recall. Notably, for k=7 and k=9, both precision and recall dropped to 0.00, resulting in an F1-score of 0.00. Figure 7 compares precision, recall and F1-score for the different values of K.

Figure 7: Precision, Recall and F1-Score

	k	Precision	Recall	F1-score
0	1	0.38	0.17	0.23
1	3	0.50	0.08	0.14
2	5	0.40	0.06	0.10
3	7	0.00	0.00	0.00
4	9	0.00	0.00	0.00



Appendix C: Code

Creating the maps

```
import folium

# for fatalities = 0 ( to create fatalities =1, change this line to 1)
data_fatality_0 = data[data['fatalities'] == 0]

# Have USA be the center
m_fatality_0 = folium.Map(location=[37.0902, -95.7129], zoom_start=4)

# adding the circles
for _, row in data_fatality_0.iterrows():
    folium.CircleMarker(location=[row['latitude'], row['longitude']], radius=3, color='blue',
fill=True, fill_color='blue').add_to(m_fatality_0)

m_fatality_0
```

Time Series plot:

```
# Group by date and count the number of demonstrations on each date
demonstrations_by_date = data.groupby('date').size()
#plotting
plt.figure(figsize=(10, 6))
plt.plot(demonstrations_by_date.index, demonstrations_by_date.values, marker='o', linestyle='-')
plt.title('Number of Demonstrations Over Time')
plt.xlabel('Date')
plt.ylabel('Number of Demonstrations')
plt.grid(True)
plt.show()
```

Protest / Year Bar Chart

```
data['year'] = data['date'].dt.year

# Group by year and count the number of protests in each year
protests_by_year = data.groupby('year').size()

plt.figure(figsize=(10, 6))
ax = protests_by_year.plot(kind='bar', color='skyblue')
plt.title('Number of Protests Each Year')
```

```
plt.xlabel('Year')
plt.ylabel('Number of Protests')
plt.xticks(rotation=45)
plt.grid(axis='y')

#count
for i, count in enumerate(protests_by_year):
    ax.text(i, count + 10, str(count), ha='center', va='bottom')
plt.tight_layout()
plt.show()
```

KNN Classification for Binary Events

```
#combining three columns into one and dropping the originals
data['date'] = pd.to_datetime(data[['year', 'month', 'day']])
data.drop(['year', 'month', 'day'], axis=1, inplace=True)

#standardize
data['latitude'] = (data['latitude'] - data['latitude'].mean()) / data['latitude'].std()
data['longitude'] = (data['longitude'] - data['longitude'].mean()) / data['longitude'].std()

#train and test data
train_data = data[data['date'].dt.year.isin([2020, 2021, 2022])]
test_data = data[data['date'].dt.year == 2023]


#distance between two points on earth's surface using longitude and latitude. Take into account earth's curvature.
def haversine_distance(lat1, lon1, lat2, lon2):
    R = 6371  # Radius of the Earth in kms
    d_lat = radians(lat2 - lat1)
    d_lon = radians(lon2 - lon1)
    a = sin(d_lat / 2) * sin(d_lat / 2) + cos(radians(lat1)) * cos(radians(lat2)) * sin(d_lon / 2) * sin(d_lon / 2)
    c = 2 * atan2(sqrt(a), sqrt(1 - a))
    distance = R * c
    return distance

#predicts test point based on majority class of its knn from training
def knn_predict(train_data, test_data, k):
    predictions = []
```

```
    for i, test_point in test_data.iterrows():
        distances = []
        for j, train_point in train_data.iterrows():
            geodist = haversine_distance(test_point['latitude'], test_point['longitude'],
train_point['latitude'], train_point['longitude'])
            distances.append((geodist, train_point['fatalities']))
        distances.sort(key=lambda x: x[0])  # Sorting based on Haversine distance
        k_nearest = distances[:k]
        counts = np.bincount([point[1] for point in k_nearest])
        prediction = np.argmax(counts)
        predictions.append(prediction)
    return predictions

#return k's accuracy
for k in range(1, 20, 2):
    predictions = knn_predict(train_data, test_data, k)
    accuracy = accuracy_score(test_data['fatalities'], predictions)
    print(f'Accuracy for k={k}: {accuracy:.2f}')
```

Precision, Recall and F1-Score

```
for k in range(1, 10, 2):
    predictions = knn_predict(train_data, test_data, k)
    true_labels = test_data['fatalities']

 class
    if any(predictions):
        precision = precision_score(true_labels, predictions, zero_division=0)
        recall = recall_score(true_labels, predictions, zero_division=0)
        f1 = f1_score(true_labels, predictions, zero_division=0)
    else:
        precision = 0
        recall = 0
        f1 = 0

    print(f'For k={k}:')
    print(f'Precision: {precision:.2f}')
    print(f'Recall: {recall:.2f}')
    print(f'F1-score: {f1:.2f}')
```