Machine Learning Prediction of Successful Rocket Launches at NASA Kennedy Using Lightning Data

Thanika Puhazholi[a,*]

[a] *Newport High School*

Abstract

This project focuses on using machine learning (ML) to predict the success of rocket launches at NASA's Kennedy Space Center by analyzing historical lightning data and assessing the significance of other weather parameters. Launch delays caused by weather conditions, particularly lightning, pose significant challenges due to their unpredictability and the high financial cost of disruptions. To address this issue, the research develops a machine learning model that leverages electric field sensor measurements to distinguish between safe and unsafe launch conditions. This approach is significant because it offers a data-driven alternative to the existing judgment-based methods, reducing unnecessary postponements while maintaining safety. The model integrates feature scaling, Support Vector Machines (SVMs), and Synthetic Minority Oversampling Technique (SMOTE) oversampling to effectively handle imbalanced data and improve prediction accuracy. The results demonstrate that electric field strength, when used as a feature in the model, allowed for precise differentiation between successful and postponed launches, ultimately helping ensure safer launch conditions. Overall, this work illustrates how combining environmental sensor data with AI/ML techniques can help assist in making critical decisions.

Keywords: Machine Learning, Rocket Launch Prediction, Lightning, Support Vector Machine, NASA Kennedy Space Center, Field Mill Data, Synthetic Minority Oversampling Technique

1. Introduction

1.1. Motivation

Rocket launches involve hundreds of interconnected systems, all of which must work perfectly within a short launch window. Weather is one of the most unpredictable variables, and lightning is among the most disruptive. A single strike can damage electronics, interfere with guidance systems, or endanger crew safety. To avoid these risks, NASA uses strict lightning rules and live weather monitoring, which are effective but can be overly cautious at times. This can lead to last-minute launch cancellations, costing millions of dollars and

*Corresponding author: thanika@outlook.com

pushing back tightly scheduled missions [1]. With launch activity increasing for both NASA and commercial partners, there is a clear need for better tools that can predict weather-related risks more precisely, helping reduce unnecessary delays while keeping safety as the top priority.

1.2. Focus on Kennedy Space Center

The Kennedy Space Center (KSC) is a natural choice for this study because of both its location and its importance to spaceflight. Positioned on Florida's east coast, KSC experiences some of the highest lightning activity in the United States - especially during the summer time, when many launches take place. As NASA's primary launch site for both crewed and uncrewed missions, even just a brief delay at KSC can disrupt multiple mission schedules. Fortunately, KSC maintains detailed lightning and weather records through a sophisticated network of electrostatic field sensors (field mills), providing a rich source of data for analysis. By focusing on KSC, this research tackles one of the most challenging launch environments, with findings that could later be adapted to other launch sites worldwide. The dataset used in this study was obtained from the KSC weather data archive [2].

1.3. Machine Learning for Launch Prediction

Machine learning makes it possible to turn decades of weather and launch data into a powerful decision-making tool. Traditional launch criteria depended heavily on fixed rules and expert judgment, which don't always capture subtle patterns in large datasets. By training machine learning models on historical lightning activity, weather conditions, and launch outcomes, it becomes possible to identify complex relationships that determine whether a launch should proceed. The result of such model is a confidence-based prediction that can complement NASA's existing guidelines, giving launch teams more nuanced information to work with. For KSC, such an approach could mean fewer avoidable delays, more efficient scheduling, and a greater reliability in launch operations overall [3].

2. Methodology / Data

2.1. Wind Shear and Rainfall

2.1.1. Exploring Rainfall, Wind Shear, and Delay Correlations

Based on preliminary research, rainfall and wind shear appeared to be straightforward weather features that could influence launch decisions. Both factors have clear physical impacts on flight dynamics, making them reasonable starting points for analysis [4]. However, when historical data for each was compared with actual launch outcomes, neither variable proved to be a reliable indicator in real-world rocket launch scenarios.

2.1.2. Sensor Measurements

Rainfall data came from KSC's network of ground-based rain gauges positioned throughout the launch complex, providing high-resolution records of precipitation rates over time. Wind shear measurements were gathered from wind profilers and upper-air balloon soundings, which track changes in wind speed and direction at multiple altitudes. Together, these instruments offer a detailed picture of weather conditions during historical launch windows.

2.1.3. Scatter Plot Analysis

To visualize potential relationships, scatter plots were created with launch dates along the x-axis and the corresponding rainfall or wind shear values on the y-axis. Each point was then categorized by outcome (either success or delay) and plotted in different colors. The goal was to determine whether a distinct threshold existed, with delayed launches clustering on one side of a dividing line and successful launches on the other.

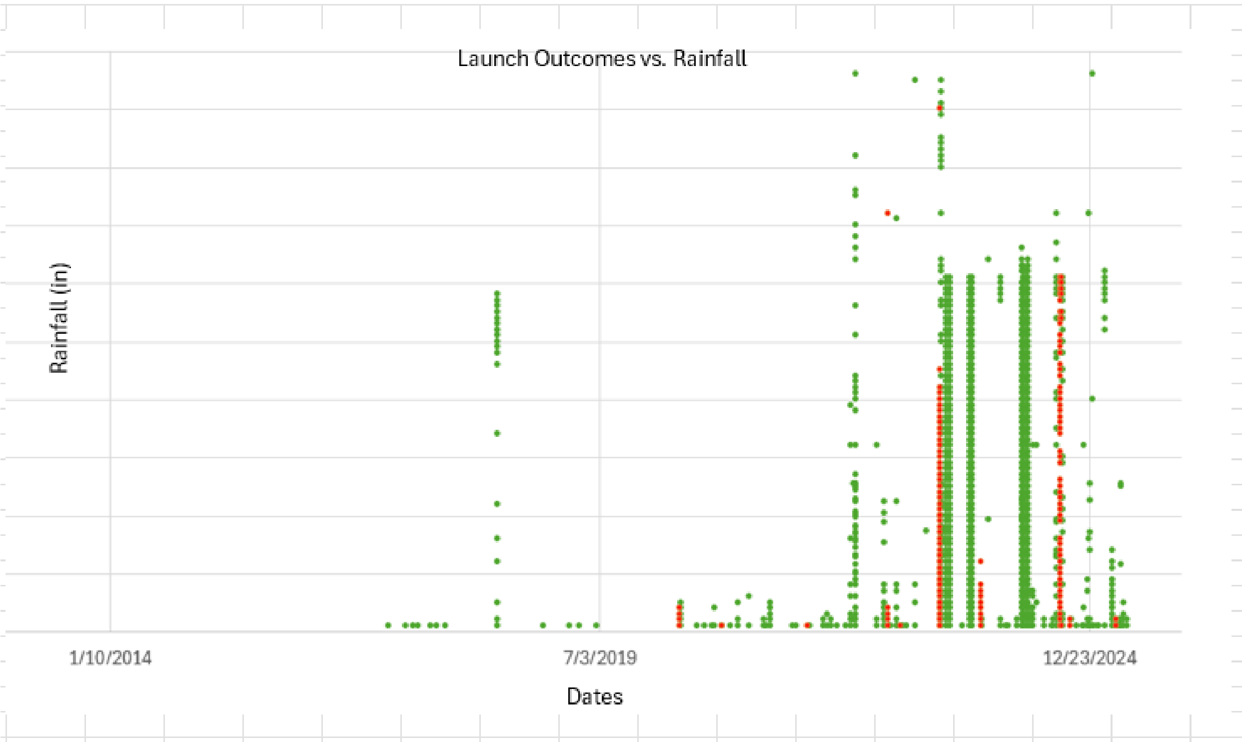

Figure 1: Launch outcome vs. rainfall; no clear pattern observed

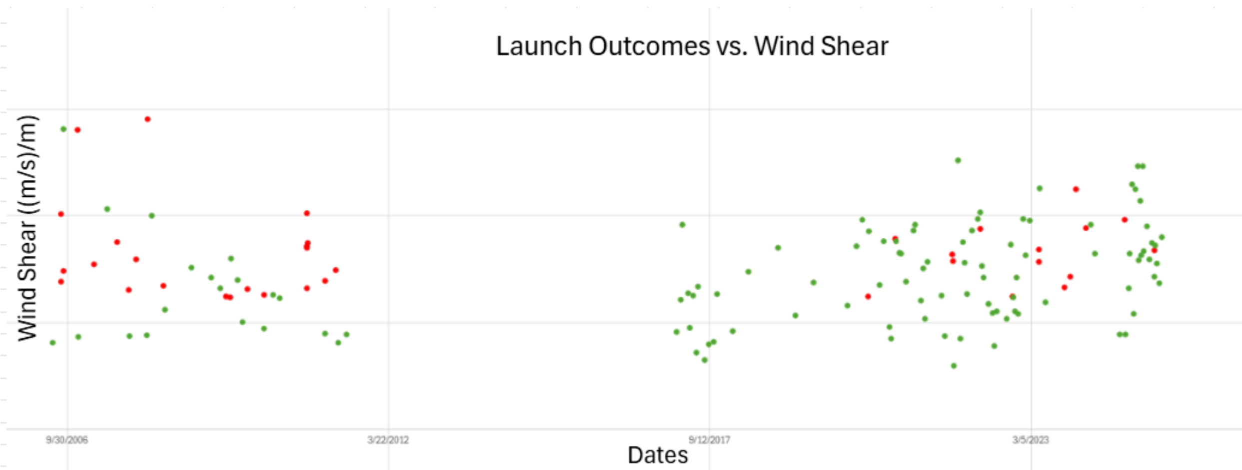

Figure 2: Launch outcome vs. wind shear; no strong correlation found

2.1.4. Lack of Correlation

The results showed no strong or consistent link between either rainfall or wind shear and launch delays. The scatter plots revealed that there was no clear dividing line separating delayed launches from successful ones - data points for both outcomes overlapped throughout the range. While certain extreme wind shear events coincided with delays, the overall trend was weak and unreliable. Rainfall showed an even less defined pattern, with measurements corresponding to both delayed and successful launches scattered across the same range.

2.1.5. Prioritizing Lightning in Launch Prediction

Discussions with meteorologists at the Kennedy Space Center Weather Office helped clarify these findings. The weather team emphasized that lightning is the most critical environmental factor affecting launch decisions, both due to its frequency in Florida and the severity of its potential impact. This guidance shifted my project's focus toward lightning analysis, which turned out to be significantly more effective for forecasting launch outcomes.

2.2. Lightning

2.2.1. Impact on Rocket Launches

Lightning poses one of the greatest risks to rocket launches, particularly at KSC, where summer thunderstorms are frequent. A strike can disrupt guidance systems, damage electrical components, or even jeopardize the rocket's structural integrity. Unlike rainfall or wind shear, which engineers can often accommodate through operational adjustments or design modifications, lightning presents an unpredictable hazard that cannot be easily mitigated. As a result, lightning remains the most critical weather parameter in launch decisions - ultimately taking precedence over other environmental factors [5].

2.2.2. Sensor Measurements

Lightning-related measurements at KSC are collected using field mills, which are clusters of electrostatic sensors that measure the local electric field in volts per meter. The data is reported as a 'one-minute mean', which means that each value represents the average electric field measured over a 60-second interval, helping to smooth out rapid fluctuations [6]. These readings provide a detailed record of how atmospheric electrical activity changes over time around the launch complex and offer an accurate measure of lightning risk, allowing for comparison with past launch outcomes.

2.2.3. *Scatter Plot Analysis*

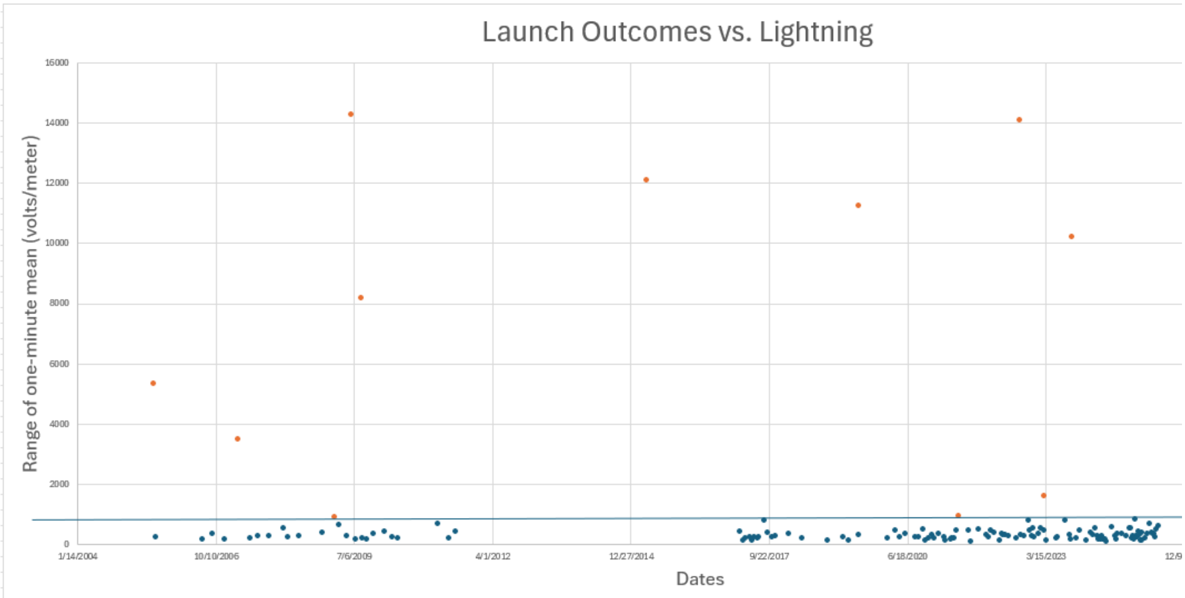

Figure 3: Launch outcome vs. lightning; clear separation by outcome

The scatter plot comparing field mill measurements with launch outcomes revealed a much more distinct pattern than those for rainfall or wind shear. Each launch date was plotted on the x-axis, with the corresponding electric field value on the y-axis, and points were color-coded by launch outcome. Unlike the previous variables, the delayed and successful launches now formed separate clusters on either side of a horizontal threshold line. This clear separation revealed the reliability of electric field strength in determining whether a launch should be successful or postponed.

2.2.4. *Threshold Selection*

From these results, I set a 60 % success probability as the threshold for launch decisions, corresponding to an electric field strength of 915 V/m. In other words, if the model predicted less than a 60 % chance of a safe launch, the final status was classified as 'postpone'. While the threshold is somewhat approximate, it aligns well with the observed data: successful launches never exceeded 839 V/m, and delayed launches never fell below 943 V/m. Setting the cutoff at 915 V/m puts it between the highest successful and the lowest delayed values, creating a buffer zone that accounts for uncertainty while still prioritizing overall launch safety.

3. Machine Learning Prediction Model

3.1. Process

3.1.1. Flow Diagram

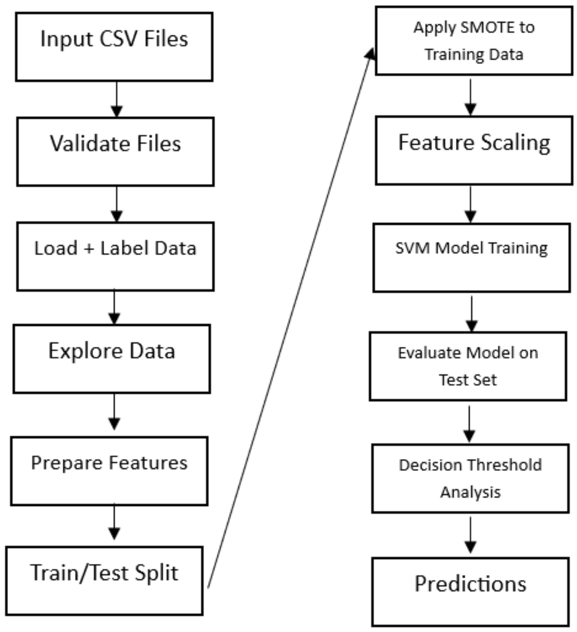

Figure 4: Overview of ML steps: preprocessing, training, classification

The analysis began by loading and labeling launch data, followed by an initial data exploration to identify patterns distinguishing successful and postponed launches.Then, key features, particularly the electric field, were prepared and divided into training and test sets (80% for training and 20% for testing). To fix class imbalances, SMOTE oversampling generated synthetic samples for the minority class. All features were subsequently scaled so the model could fairly compare them (preventing larger numbers from dominating the smaller ones). Next, an SVM model was trained, learning how to separate successful from postponed launches by finding the boundary that best splits the data. Finally, the model's performance was evaluated on the test set, a decision threshold was identified, and predictions were produced to estimate launch success probabilities.

3.1.2. Underlying Background Equations

Two primary equations formulate the mode. First, feature scaling standardizes each variable using:

$$x_{\text{scaled}} = \frac{x - \mu}{\sigma} \tag{1}$$

where μ is the mean and σ is the standard deviation of the feature. This ensures that all features are on the same scale, preventing larger values from dominating the model [7].

Second, the SVM separates classes using a hyperplane:

$$f(x) = w \cdot x + b \tag{2}$$

Here, x represents the feature (electric field), w is the weight that determines the slope, and b is the intercept. A launch is classified as successful if $f(x) > 0$ and classified as postponed if $f(x) < 0$. In other words, the hyperplane acts like a boundary dividing the two outcomes [8]. To estimate probabilities, the output is passed through a sigmoid function:

$$P(\text{ success }) = \frac{1}{1 + e^{-f(x)}} \tag{3}$$

This converts the SVM's raw output (the distance of the data point x from the hyperplane) into a value between 0 and 1, representing the likelihood of a launch being successful. A higher $f(x)$ results in a probability closer to 1, while a lower $f(x)$ gives a probability closer to 0 [9].

3.2. k-Nearest Neighbors (KNN)

3.2.1. Data Imbalance and Oversampling

The original dataset contained 135 successful launches but only 10 postponed launches, creating a highly imbalanced situation. As a result, the first model was heavily biased toward predicting success, often classifying launches as successful even in ranges where the scatter plot clearly showed they should have been postponed. This imbalance made the model unreliable for safety-critical decisions. To address this, SMOTE oversampling was applied to generate 125 synthetic postponed launches, creating a balanced dataset with 135 successful and 135 postponed points. SMOTE works by creating new and realistic data points between existing ones. For example, if postponed launches exist at electric fields of 900 and 1000, SMOTE might generate synthetic points at 925, 950, or 975.

3.2.2. Underlying Background Equations

SMOTE generates synthetic points using the equation:

$$\mathbf{x}_{\text{new}} = \mathbf{x}_i + r \cdot (\mathbf{x}_{\text{neighbor}} - \mathbf{x}_i) \tag{4}$$

Here, x_i is an existing data point from the minority class, x_{neighbor} is one of its nearest neighbors, and r is a random number between 0 and 1. Essentially, this creates new points along a line connecting the real values - enriching the minority class and ultimately allowing the model to better recognize patterns that distinguish postponed launches from successful ones [10].

4. Results and Discussion

4.1. Example Run

```
Enter magnetic range (115.0-14302.0): 250
📊 Prediction for Range 250.00 (SMOTE Model):
   Status: Successful
   Success Probability: 95.0%
   Confidence: Very High (95.0%)

Enter magnetic range (115.0-14302.0): 800
📊 Prediction for Range 800.00 (SMOTE Model):
   Status: Successful
   Success Probability: 69.9%
   Confidence: Medium (69.9%)

Enter magnetic range (115.0-14302.0): 915
📊 Prediction for Range 915.00 (SMOTE Model):
   Status: Successful
   Success Probability: 60.0%
   Confidence: Medium (60.0%)

Enter magnetic range (115.0-14302.0): 2000
📊 Prediction for Range 2000.00 (SMOTE Model):
   Status: Postponed
   Success Probability: 2.4%
   Confidence: Very High (97.6%)

Enter magnetic range (115.0-14302.0): 14000
📊 Prediction for Range 14000.00 (SMOTE Model):
   Status: Postponed
   Success Probability: 0.0%
   Confidence: Very High (100.0%)
```

Figure 5: Model prediction output showing launch success probability

4.2. Implications and Potential Uses

The developed model can function as a decision support tool at the Kennedy Space Center by providing a clear, data-driven assessment of when launch conditions are safe versus when they present significant risk. In doing so, it reduces the reliance on subjective human judgment, helping to limit unnecessary postponements and unsafe launches that carry substantial hazards. Beyond launches, this approach highlights how machine learning can translate raw weather measurements into meaningful insight, suggesting potential applications in other fields where safety depends on real-time environmental data.

5. Conclusions

This study developed a machine learning system capable of predicting whether rocket launches at the Kennedy Space Center should proceed or be postponed, based on historical lightning data. By integrating feature scaling, SVM, and SMOTE oversampling, the model effectively handled imbalanced data and learned to distinguish safe from unsafe launch conditions. The results demonstrate that electric field measurements provide a reliable indicator of launch risk, providing accurate predictions that limit avoidable delays without compromising safety. Overall, this work illustrates how environmental sensor data can guide critical decisions with confidence through the use of AI.

References

[1] R. Miteva, S. W. Samwel, S. Tkatchova, Space weather effects on satellites, Astronomy 2 (3) (2023) 165–179. doi:10.3390/astronomy2030012.
URL https://www.mdpi.com/2674-0346/2/3/12

[2] K. Smith, John F. Kennedy Spaceport Weather Archive, https://kscweather.ksc.nasa.gov/wxarchive, [Online; accessed 17-Sep-2025] (2025).

[3] NASA, Space shuttle weather launch commit criteria and ksc end of mission weather landing criteria, Tech. rep., National Aeronautics and Space Administration (2007).

[4] L. Aldaghma, D. Muresan, S. Renaud, Establishing the requirements for safe rocket launches with respect to weather, Acta Astronautica 213 (2023) 392–407. doi:https://doi.org/10.1016/j.actaastro.2023.07.008.
URL https://www.sciencedirect.com/science/article/pii/S0094576523003594

[5] S. Gardner, E. White, B. Langhals, T. McNamara, W. Roeder, A. E. T. Jr, A field-mill proxy climatology for the lightning launch commit criteria at cape canaveral air force station and nasa kennedy space center (2024). arXiv:2403.07016.
URL https://arxiv.org/abs/2403.07016

[6] K. Smith, John F. Kennedy Spaceport Weather Archive, https://kscweather.ksc.nasa.gov/wxarchive, [Online; accessed 5-Oct-2025] (2025).

[7] J. M. H. Pinheiro, S. V. B. de Oliveira, T. H. S. Silva, P. A. R. Saraiva, E. F. de Souza, R. V. Godoy, L. A. Ambrosio, M. Becker, The impact of feature scaling in machine learning: Effects on regression and classification tasks (2025). arXiv:2506.08274.
URL https://arxiv.org/abs/2506.08274

[8] S. yin Xia, Z. yang Xiong, Y. guo Luo, L. mei Dong, A method to improve support vector machine based on distance to hyperplane, Optik 126 (20) (2015) 2405–2410. doi:https://doi.org/10.1016/j.ijleo.2015.06.010.
URL https://www.sciencedirect.com/science/article/pii/S0030402615004829

[9] S. R. Mugunthan, D. T. Vijayakumar, Design of improved version of sigmoidal function with biases for classification task in elm domain, Journal of Soft Computing Paradigm 03 (02) (2021) 70–82. doi:https://doi.org/10.36548/jscp.2021.2.002.
URL https://web.archive.org/web/20210611052941id_/https://irojournals.com/jscp/V3/I2/02.pdf

[10] C. Gong, L. Gu, A novel smote-based classification approach to online data imbalance problem, Mathematical Problems in Engineering 2016 (1) (2016) 5685970. arXiv:https://onlinelibrary.wiley.

com/doi/pdf/10.1155/2016/5685970, doi:https://doi.org/10.1155/2016/5685970.
URL https://onlinelibrary.wiley.com/doi/abs/10.1155/2016/5685970