Evaluating the Effect of "Super Spikes" on Distance Running Performance

Brendan Cave

bcave1@umbc.edu

I. Introduction

In 2020, Nike released two models of track spikes which revolutionized distance running across all levels: the ZoomX Dragonfly and the Air Zoom Victory. That same year, Ugandan athlete Joshua Cheptegei broke the world records in both the 5,000m and 10,000m run while wearing the Dragonfly spikes. The following months saw numerous other records across high school, college, and professional track and field fall, prompting followers of the sport to label the Dragonfly and Victory as "super spikes."

Blaming the spikes for the explosion of elite performances was not an unreasonable conclusion. These shoes featured extreme technological advancements in the form of Nike's ZoomX foam and carbon fiber plates, first seen in the Vaporfly 4%. ZoomX foam is both extremely lightweight and resilient, meaning that much of the energy that a runner exerts on the ground is returned to them as they stride forward. The carbon fiber plate, which sits below the foot, further saves energy by flexing and rebounding as the foot strikes and leaves the ground. These factors come together in the Dragonfly and Victory to reduce the amount of energy expended by a runner during each step, producing large benefits over a distance race in which a runner takes thousands of steps.

Photo of the Dragonfly. The plate is sandwiched between layers of green ZoomX foam.

Image: https://www.soccerplususa.com/prodimages/32578-BLACKLIME_BLAST-l.jpg

The success of super spikes both on the track and in stores inevitably led to the question of fairness. If the spikes provide enough of a benefit to enhance performance, is it fair to allow their use in competitions? Athletes unable to access this technology would face a disadvantage, and furthermore, comparing times set in super spikes to records set without them would be similarly unfair. The other question is one faced by athletes: are super spikes worth it? The Dragonfly costs $160 with the Victory at an even steeper $190, a seemingly excessive price when traditional spikes can be found for less than $75. However, with the competitiveness of the sport, missing out on potential performance-enhancing technology could cost somebody precious seconds or even a victory.

As a collegiate distance runner, I wanted to answer for myself the question of whether super spikes truly did improve athletes' performances. I decided to do so using machine learning algorithms, specifically those that classify data. I selected multiple instances of the NCAA Outdoor Track and Field Championships, some before the release of super spikes and some after. I then fit multiple machine learning algorithms to the data using Jupyter Notebook and assessed the algorithms' ability to classify an athlete's performance as one run with or without super spikes. In general, the algorithms managed to accurately select whether times were run before or after super spikes, so I concluded that this technology had produced an improvement in athlete performances.

II. Related Work

Not much work has been done in assessing the performance benefits offered by super spikes, and I have not found any studies that use a machine learning approach. One relevant study from The National Center for Biotechnology Information is an article titled, "Can We Quantify the Benefits of 'Super Spikes' in Track Running?" This did not delve into whether the shoes caused athletes to run faster, but it offered metrics for evaluating the specific enhancements offered by the technologies available in super spikes. In its conclusion, the article suggested that "rather than relying on lab-measured predictors of track running performance, we might just need to rely on comparison of track performances pre and post the introduction of super spikes." This is exactly the approach I decided to use in my study.

III. Proposed Method

In order to examine whether super spikes have enhanced running performance, I decided to assess how well four machine learning algorithms performed in classifying an athlete's performance as one run before or after the release of super spikes. I selected k-NN, perceptron, Naïve Bayes, and logistic regression to classify the data, which are algorithms commonly used in classification problems. I will briefly describe how each algorithm works.

k-NN uses the k closest points to the point of interest in order to classify the point of interest. k is selected prior to fitting the algorithm. Whichever label is the most common among the k nearest neighbors to the test point is applied to the test point.

The perceptron algorithm attempts to separate the classes in the dataset using a hyperplane. The hyperplane equation is as follows: $y = w^T x + b$, where y is the label of the point x. w and b are a set of weights and biases which are fit to the dataset. During training, each point is compared to the hyperplane by inserting it into the equation $y = w^T x + b$. If it is correctly classified, the hyperplane does not change, but if it is misclassified, the hyperplane is updated using the equation $w_{new} = w_{old} + xy$. Some datasets are not linearly separable, so in these cases, a maximum number of iterations must be specified so the algorithm does not run forever attempting to separate the data.

Naïve Bayes uses probability to classify a point based on its features. It assumes that the predictors each have an equal impact on the outcome. I used a Gaussian Naïve Bayes model because the predictors in this case are continuous, as they are times. The distribution of times run also appeared somewhat normal, so a Gaussian distribution fit the features well. One limitation of the Naïve Bayes classifier in this particular case is that it assumes all predictors are independent, which is not entirely true. My predictors are overall time and last lap time, which can impact each other. A faster race generally leaves athletes more fatigued by the end and produces slower last lap times, while a slower race has the opposite effect.

Logistic regression uses the predictor variables to estimate the probability of the outcome event. A log transformation is applied on the odds, or probability of success divided by probability of failure, producing a situation in which a label is applied based on whether the probability is greater than or less than 0.5.

I used these four algorithms on my dataset to observe how accurately they classified each point as one run before or after super spikes. If they performed particularly well, then I would conclude that super spikes have caused a difference in distance running performance.

IV. Experiments

Dataset: I produced the dataset in this experiment myself. The NCAA has results from each year's track championships available online at ncaa.com, and the results include splits for each lap. Since Nike released the first super spikes to the public in 2020, I decided to use an equal number of championships before and after 2020. I chose the 2017, 2018, and 2019 championships as the pre-super spikes era and the 2021, 2022, and 2023 championships as the super spikes era. No championship event occurred in 2020 due to the COVID-19 pandemic. Using results from the same annual event would reduce the influence of external factors, as the athletes were relatively the same age and competed at the same level. This allowed me to better isolate the effect of the shoes.

I decided to use the 5,000m run from the NCAA Championships because super spikes display greater benefits over long distance races. I opted not to use the 10,000m run because temperature and humidity can greatly impact an event as long as the 10,000m- high heat and humidity force athletes to run more slowly to avoid overheating.

My dataset included four variables- year, SS, time, and last lap. Year represented the year that the race occurred. SS was an indicator variable that equaled 0 if the race occurred before the release of super spikes and 1 otherwise. Time was the total time it took the athlete to cover 5,000 meters, and last lap is the time it took to cover the final 400-meter lap of the race.

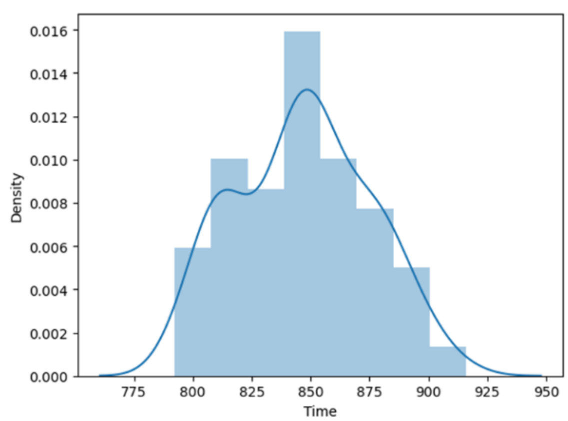
	Year	SS	Time	Last Lap
0	2017	0	875.60	55.76
1	2017	0	875.88	56.19
2	2017	0	876.23	56.41
3	2017	0	876.57	56.31
4	2017	0	876.78	56.69

The first five entries of my dataset

Software: I used the Python programming language to make computations, which I accessed through Jupyter Notebook. I also used a variety of libraries, including Pandas, Numpy, Matplotlib, Seaborn, and Scikit Learn. I used these resources to explore my dataset and make visualizations about it. I then fit models for each machine learning algorithm to my dataset and tested their ability to classify test points.

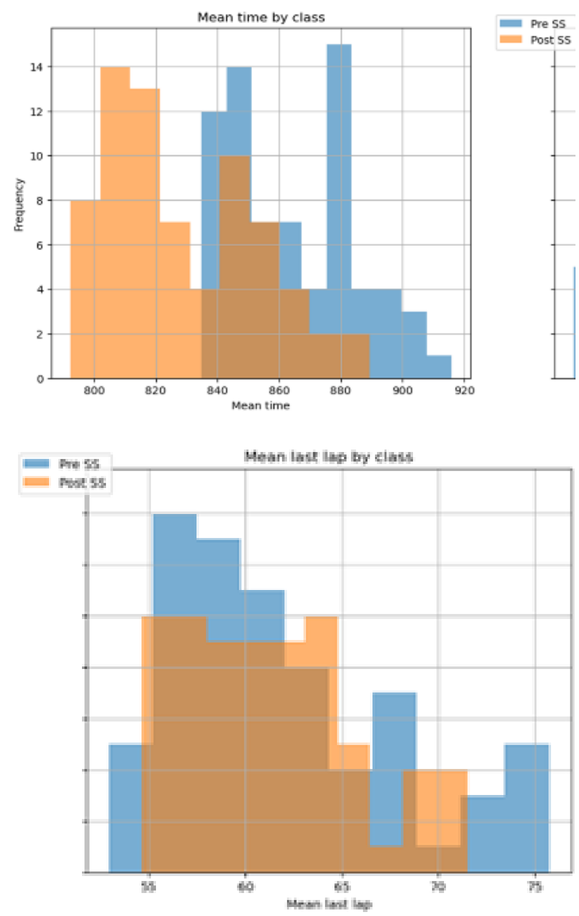
V. Results and Discussion

<u>Dataset Exploration:</u> I first observed the distribution and summary statistics of the overall dataset. The 5,000m times of the athletes were distributed somewhat symmetrically around the mean of 846.74 seconds (14:06.74). The median of 848.02 (14:08.02) was close to the mean, implying little skewness. The standard deviation was 28.52 seconds. The last lap times also displayed little disparity between the mean and median- the mean was 61.59 seconds and the median was 60.93.



The distribution of times for all athletes in the dataset

I then separated the data by class and graphed the distributions to observe possible differences. The times from the super spike era appeared faster in general than those from the pre-super spike era. There was some overlap, but a majority of times run with super spikes were faster than all times run without super spikes. The distributions of last lap times did not differ much between the pre- and post- super spike eras, indicating that super spikes did not have much influence on last lap times.



The distributions of time and last lap separated by SS

<u>Machine Learning Algorithms:</u> I separated the data into 90% training and 10% testing data and used this split for each model. This eliminated any possible disparities between the models based on using different data for fitting and testing.

<u>k-NN:</u> I first selected a value for k using cross-validation. I set k equal to integers from 1 to 10 and used 10-fold cross validation to evaluate the accuracy of each model. This process involved separating the data into 10 groups, with each serving as the test set once for each number of neighbors. The best-performing model used $k = 2$, which returned an accuracy score of 0.85. After receiving feedback on this result, I decided to not use $k = 2$ because this allows for ties- if the two nearest neighbors to a test

point are of different classes, then the label must be assigned randomly. Instead, I selected 3 for k because this returned the second-best accuracy score of 0.814. When I fit a k = 3 model to the training dataset, it produced an accuracy of 0.87 on the testing set.

Perceptron: I fit a perceptron model to the training dataset using 500 iterations, but when I tested it, it only produced an accuracy of 0.53. This number was rather low, so I increased the number of iterations to 1000 to see if the model could separate the data any better. The score did not change, and when I increased to 10,000 iterations, it still did not change. This implies that the classes were not linearly separable as the algorithm could not successfully classify all the points.

Naïve Bayes: I fit a Gaussian Naïve Bayes model to the training set and it produced the following parameters.

Parameter	Time	Last Lap
Mean SS=0	14:23.998	62.027
Mean SS=1	13:50.475	61.416
Var SS=0	20.992	7.837
Var SS=1	24.513	4.553

The parameters of the Naïve Bayes model

The means were very different between the pre-super spike era and the super spike era- the difference was 33 seconds. The model produced an accuracy score of 0.74 on the training set and 0.87 on the testing set.

Logistic Regression: The logistic regression model produced an accuracy score of 0.67 in training and 0.87 in testing. Like the Naïve Bayes model, it generalized well to the testing data. The fact that both models performed better in testing than training implies that neither model overfit the data. Overfitting occurs when a model fits too closely to the training dataset, causing it to perform poorly in testing. Since both models did well in testing, I concluded that neither overfit the data.

Comparison of the Models:

Model	Accuracy
k-NN	0.87
Perceptron	0.53
Naïve Bayes	0.87
Logistic Regression	0.87

Each model's accuracy score

Apart from the perceptron model, each machine learning algorithm was 87% accurate in classifying times run prior to super spikes and times run after their release.

VI. Conclusions

In general, the machine learning algorithms that I selected performed well in classifying whether a time had been run with super spikes or not. Each model, apart from perceptron, was 87% accurate with its classifications. This led me to conclude that super spikes have enhanced distance runners' performances and allowed them to achieve faster times. However, this is not a decisive conclusion. I recommend further research into athletes' performances at other levels of the sport, such as high school and professional running. If the results are similar at other levels, the argument that super spikes have produced faster times becomes more convincing. Also, super spikes have only been in use for three years. Distance running results should be monitored into the future to see if the trend of faster times continues, or if it was a short-term phenomenon. Observing future results will better isolate the impact of super spikes and reduce the influence of confounding variables.

VII. Limitations

I worked on this project myself, so instead of a Contributions section, I will call attention to some factors that limit the decisiveness of my conclusion. First, each NCAA Championship features a different roster of athletes. Some manage to qualify for the event multiple times, but much of the field differs from year-to-year. Each athlete possesses different abilities and follows a different training program, influencing the times they run. Also, the location of the NCAA meet changes each year. As mentioned earlier, weather conditions impact distance races such as the 5,000m. The difference is not as drastic as in the 10,000m, but it still makes a difference. For example, an athlete ran 13:12 to win the 2021 meet, which featured cool 63F temperatures. Meanwhile, in the 2023 championships, a much hotter 94F day caused the winning time to slip to 14:04, nearly a minute slower. Differences in finishing times could be attributed to these factors and not super spikes, reducing the decisiveness of the conclusion. One method proposed to me to improve this study was to use the NCAA Indoor championships, which would negate the impact of weather as indoor temperatures are controlled. I would be interested to see if the conclusion holds, and if times vary less from year-to-year.

VIII. Acknowledgements

This project was made possible by the NCAA for making the results from each championship meet publicly available online and including splits for each individual lap. It was also aided by Dr. Thu Nguyen, who taught me each of the machine learning algorithms I used and recommended selecting an odd number for k to prevent ties.

IX. References

Gandhi, Rohith. "Naïve Bayes Classifier." *Medium.* 5 May 2018. https://towardsdatas cience .com/naive-bayes-classifier-81d512f50a7c. Accessed 16 Dec 2023.

Healey L, Bertschy M, Kipp S, Hoogkamer W. "Can We Quantify the Benefits of "Super Spikes" in Track Running?" *Sports Med.* 2022 Jun; 52(6):1211-1218. doi: 10.1007/s40279-022-01657-4. Epub 2022 Feb 23. PMID: 35195880; PMCID: PMC8864588. Accessed 15 Dec 2023.

Hutchinson, Alex. "How do Nike's Vaporfly 4% Shoes Actually Work?" *Outside.* 20 Nov 2018. https://www.o utsideonline.com/run ning/how-do-nikes-vap orfly-4-shoes-actually-work/?scope=anon. Accessed 17 Dec 2023.

"What is Logistic Regression?" *IBM.* https://www.ibm.com/topics/logistic-regression. Accessed 18 Dec 2023.

https://www.ncaa.com/sites/default/files/external/track-field/results/d1/outdoor17/final/007-1-01.htm

https://www.ncaa.com/sites/default/files/external/track-field/results/d1/outdoor18/Finals/007-1_compiled.htm

https://dt8v5llb2dwhs.cloudfront.net/NCAA/007-1-01.htm

https://dt8v5llb2dwhs.cloudfront.net/Outdoor/2021/007-1-01.htm

https://flashresults.ncaa.com/Outdoor/2022/007-1-01.htm

https://flashresults.ncaa.com/Outdoor/2023/007-1-01.htm