DS2500 - Final Project Report Analysis

Dan Beecher (beecher.d@northeastern.edu), Rayd Hussain (hussain.r@northeastern.edu)

Problem Statement and Background

Dating back all the way to the 9th century, scientists have since been recording meteorite landings to gain insight into the composition of the material, as well as their interaction with the Earth's terrain. This project aims to analyze the distribution patterns of meteorite landings across Earth's surface, investigating the complex interplay of factors such as geographical features, atmospheric conditions, and preservation biases that may influence where these landings ultimately die down. Our motivation stems from the opportunity to connect celestial phenomena with terrestrial consequences, as meteorites serve as natural probes from space, helping identify crucial information about our solar system's formation and evolution. The global scientific community, including NASA, the Meteoritical Society, and many other research institutions worldwide, has long recognized the value of meteorites and has been maintaining detailed records combined with technological advances in observation networks and machine learning algorithms, that help enable more sophisticated analysis than ever before. The significance of this research extends beyond academic interest to areas including public safety, resource identification, and potential contributions to broader planetary defense strategies. An implication of applying AI to landings across the world and especially the U.S. would begin the accelerated development of adaptive regional prediction systems that could identify previously undetected geographical and temporal patterns across different world regions. These systems could generate highly localized prediction models customized to the characteristics of diverse world ecosystems. This approach could reveal how different meteorite types survive in specific regions, enable authorities to establish region-specific recovery protocols, and create an intelligent early warning network that continuously improves with the feeding of new data, ultimately transforming meteorite science from general global predictions to targeted regional forecasting with significantly higher accuracy rates.

DS2500 - Final Project Report Analysis

Dan Beecher (beecher.d@northeastern.edu), Rayd Hussain (hussain.r@northeastern.edu)

Introduction to our Data

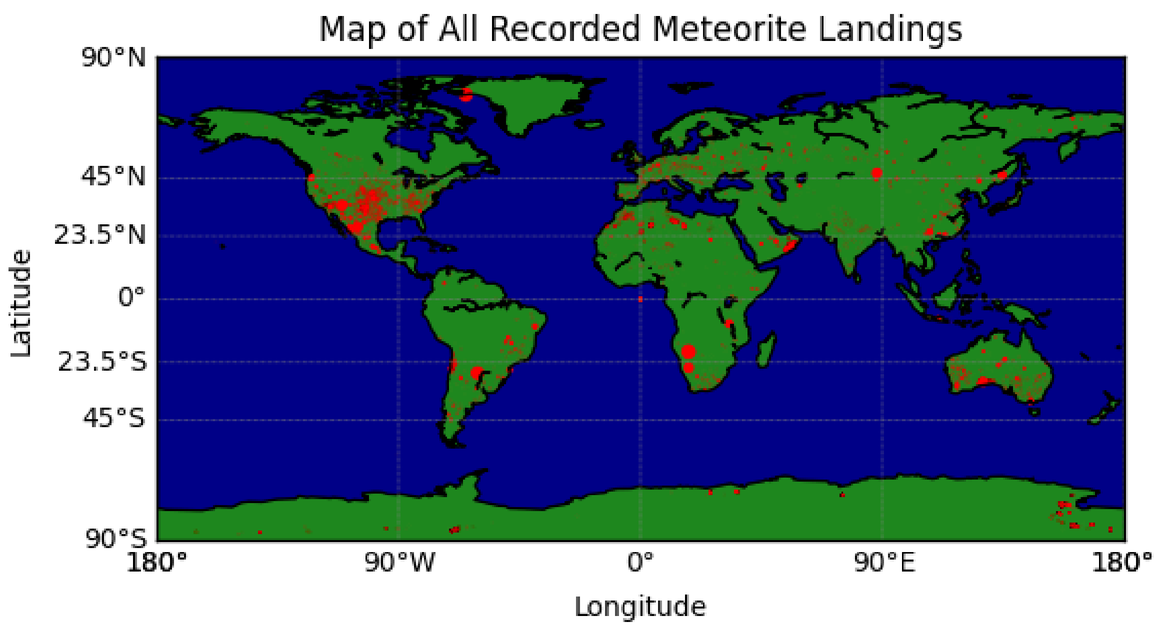
The dataset is titled the "Meteorite Landings", which compiles records of meteorite impacts on Earth. It originates from public data sources managed by the U.S. government and scientific institutions, most notably NASA, and is available through the data.gov catalog. Being recently updated as of April of 2025, this ensures all data is accurate and listed in real-time. The records were collected through a combination of field observations, museum records, scientific research, and historical accounts. Meteorite data is generally gathered by research teams and collectors who report their findings to centralized databases managed by organizations like NASA. There are no significant privacy or ethical concerns associated with this dataset because it contains no personal or sensitive information since its focus is solely on natural phenomena. Bias may be present in this dataset due to several factors. For example, meteorite recoveries are more likely to occur in heavily populated areas where observations are frequent, as opposed to remote or uninhabited regions where landings might go unnoticed. Additionally, preservation conditions and reporting efforts can vary widely between regions, which may skew the geographic distribution of recorded meteorite landings. This indicates that while the dataset is invaluable for studying meteorite impacts, it should be interpreted with an understanding that under-reporting or over-reporting in certain regions could lead to potential biases in the analysis.

Data Science Approaches

The analysis employs a combination of data analysis, geospatial visualization, and predictive modeling techniques, initially, data normalization is used to standardize the meteorite mass values, ensuring that the feature scales do not bias subsequent analysis. Geospatial techniques are then applied to visualize the distribution of meteorite landings using mapping tools. These visualizations not only illustrate the overall distribution across the globe and specifically within the United States but also highlight temporal trends by segregating the landings into distinct historical periods. For the predictive component, a k-nearest neighbors (KNN) classification model is implemented. This algorithm predicts the meteorite class based on features such as mass and geographic coordinates (longitude and latitude). This dataset is divided into training and testing subsets to evaluate the model's performance, and various values of k are tested to identify the most accurate parameter setting. This approach helps in understanding how well the meteorite classification can be predicted based on the extracted features, thereby offering insights into the underlying patterns in the meteorite landing data.

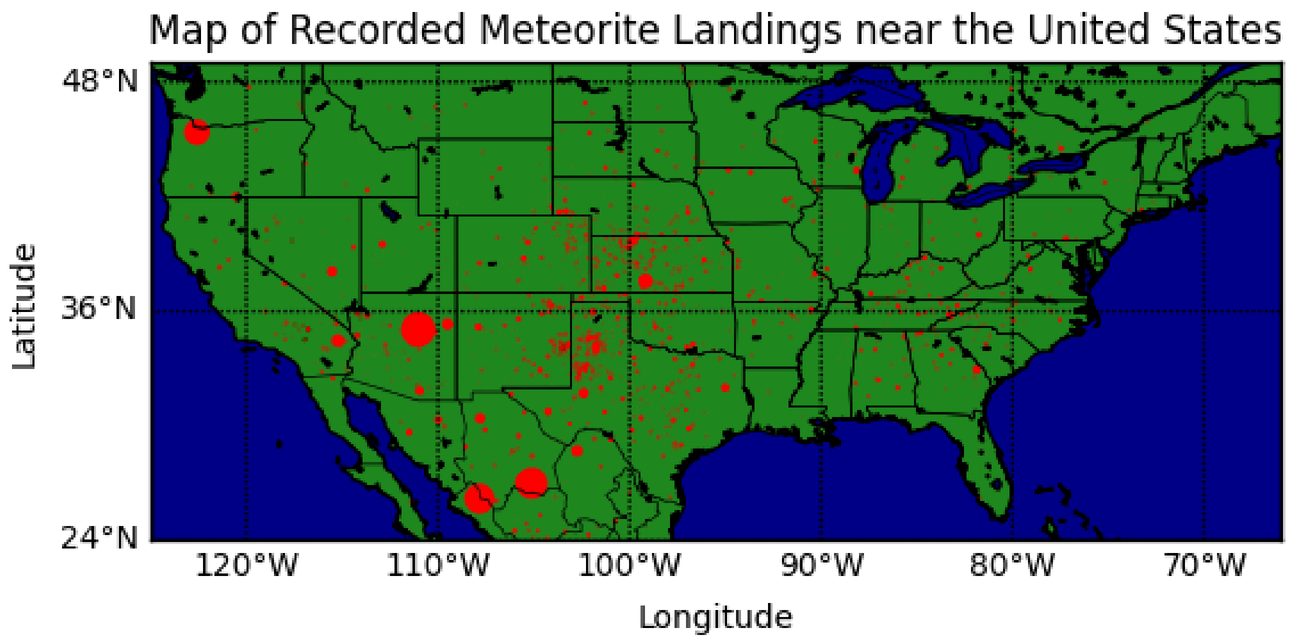
Results and Conclusions

This in-depth analysis produces a total of four visualizations, with the first being a world map of all recorded meteorite landings contained within the dataset:



Python's Basemap library is used to plot the initial world map and the meteorite landing locations are layered on top of the map. The geometric data contained within the dataset for each meteorite landing consists of the latitude and longitude of the recorded landing position, making the process of layering the data on the graph pretty simple. The size of each dot is indicative of the normalized mass of each meteorite, with the scale being increased to improve the appearance and visibility. An interesting aspect of this visualization is the disproportionate amount of meteorites that landed in the United States. A potential explanation for this could be attributed to the dataset originating from data.gov, which is the United States government's open data website. It is plausible that it was easier to gather data about meteorite landings in the United States as opposed to other outside territories, implying that this dataset likely doesn't contain every meteorite landing that has occurred since meteorite landing data first started being collected.

The second visualization zoomed in on the continental United States and the surrounding area,

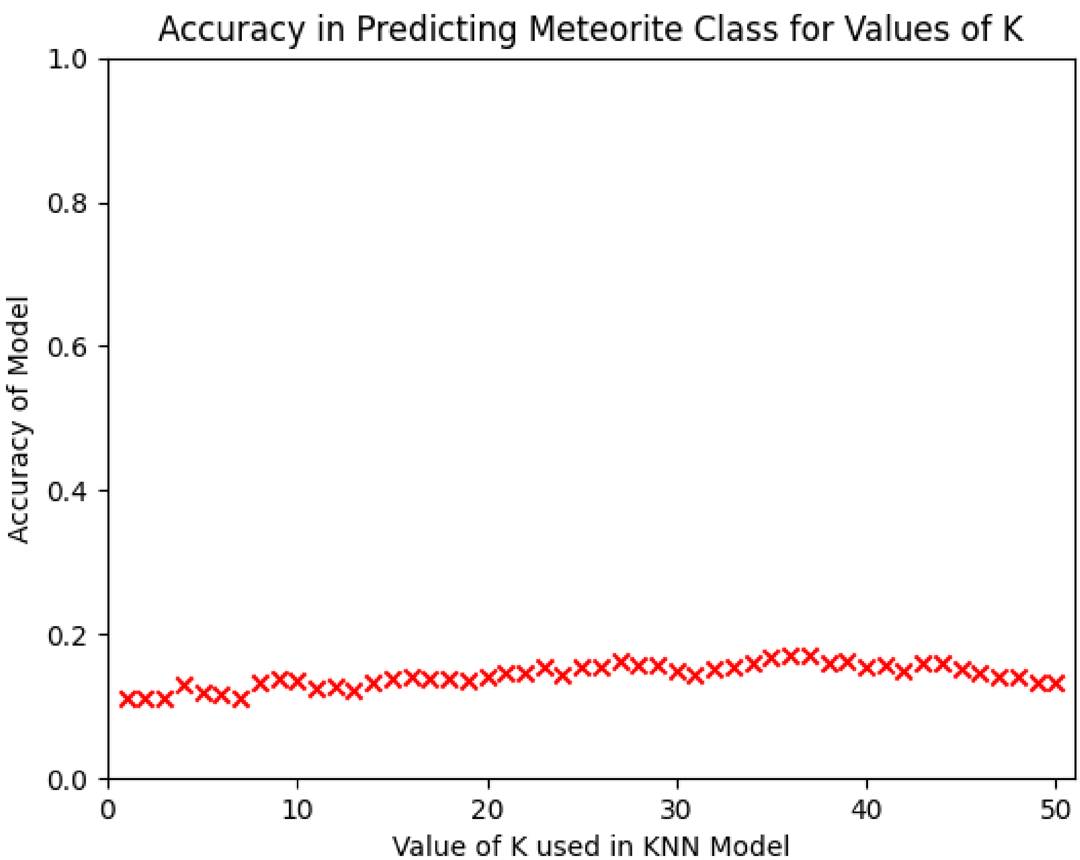


allowing for a better look at the meteorite landings there. Basemap is used again, but the longitude and latitude parameters are adjusted to zoom in on this area. The meteorite landing dataset is narrowed down to just the meteorites that landed within the same bounds as the map. Once again, the size of the plotted points are representative of the mass of the meteorite, but the scalar is adjusted from the first visualization to improve visibility. This visualization reveals that many of the meteorites that landed in this area occurred around 110 to 100 degrees west longitude and that minimal meteorites were recorded to have landed east of 80 degrees west longitude. Geological characteristics of the Midwest could explain the high concentration of meteorite landings in this region, but more research will have to be done to find the true explanation.

DS2500 - Final Project Report Analysis

Dan Beecher (beecher.d@northeastern.edu), Rayd Hussain (hussain.r@northeastern.edu)

Another aspect of the analysis includes a k-nearest neighbors classification model, where the mass and landing location are used to predict the class of each meteor. The U.S. area meteorite landings are sorted into training and testing sets and a function finds the accuracy for values of k from 1 to 50:

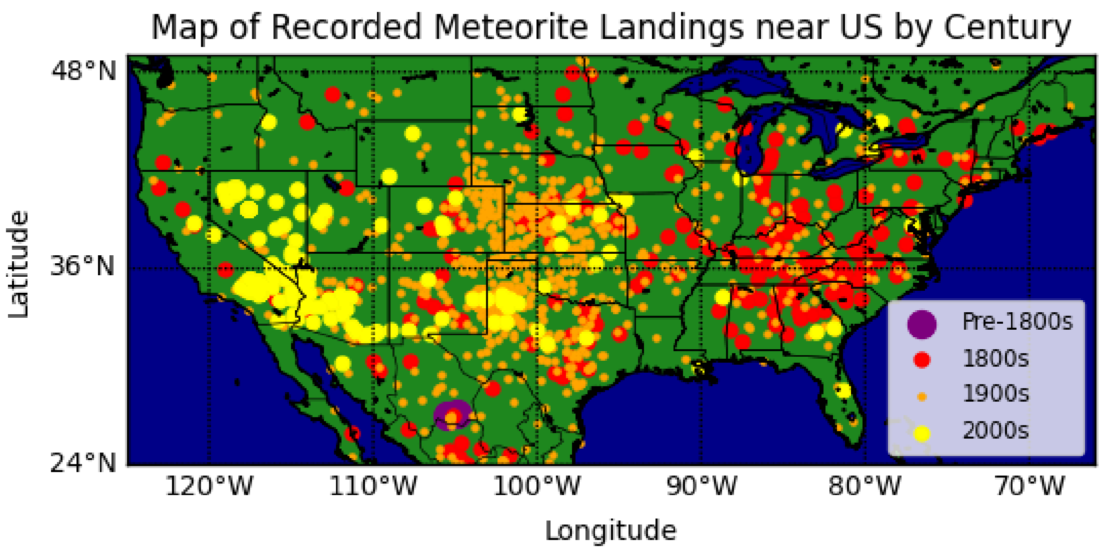


This visualization clearly shows that regardless of the K value used, the mass and landing location of meteorites is not effective at predicting the class of meteorites. There is a low variance in accuracy when looking at the different values of K used in the model, with the range of accuracies being around 10-18%. This is a pretty acceptable finding as there are over 100 different classes of meteorites amongst the meteorites that landed in or around the continental United States, and the mass and landing location of the meteorites is not very much information to use to make a prediction model. If more information was available, such as the time of year that the meteorites landed or the velocity of the meteorite upon impact, a K-nearest neighbors prediction model might be more accurate at predicting the meteorite class.

DS2500 - Final Project Report Analysis

Dan Beecher (beecher.d@northeastern.edu), Rayd Hussain (hussain.r@northeastern.edu)

The final visualization displays landings of meteorites in the United States area in different centuries. Unlike the previous two visualizations that show meteorite landings, the size is not representative of the mass of the meteorite, but is instead adjusted based on the number of meteorites from that century so that the centuries with less recorded meteorite landings stand out on the graph.



There are some interesting trends around the location of the meteors from the different centuries. An interesting trend is the proximity of recorded meteorite landings in each century. Only two meteors from before the year 1800 were recorded, and both landed in close proximity to the other in Northern Mexico. Many of the meteorites that landed in the 1800s landed East of the Mississippi River. A large proportion of meteorites that landed in the 1900s fell in the Midwest, and many of the 2000s meteorites landed in Nevada, Southern California, and Arizona. A reasonable takeaway is that there is a connection between time and location of meteorite landings.

Future Work

Since the data has relatively low information of the landings, researching and testing other datasets, especially for ones that include velocity, composition, and angle of entry/impact could refine the model's predictive capabilities and provide deeper insights into the physical processes driving these impacts. Advanced machine learning techniques, such as deep neural networks or ensemble methods, could be implemented to capture nonlinear relationships in the data, potentially increasing classification accuracy and uncovering hidden patterns. Integrating geospatial analysis with satellite imagery and environmental data could help contextualize meteorite landings within broader geographic and climatic trends. This approach might reveal correlations between meteorite impacts and local geological features or atmospheric conditions, which can help contribute to better analytical findings. Finally, developing a real-time monitoring and prediction system based on updated and reformed meteorite landing data would be a valuable extension, further supporting applications in planetary defense and disaster preparation.

DS2500 - Final Project Report Analysis

Dan Beecher (beecher.d@northeastern.edu), Rayd Hussain (hussain.r@northeastern.edu)

Reference List:

The Meteoritical Society. (2013). Meteorite Landings [Data set]. Data.gov. https://catalog.data.gov/dataset/meteorite-landings