Group Project Proposal
Meteorite Data Mine

Ziad Arafat, Kevin Dhanapal, Jason Ivey, and Jacob Yoder

Abstract—

Index Terms—Data Mining, Machine Learning, Meteorites, Proposal

1 Introduction

WE are taking up a group project as a challenge to demonstrate the KDD framework. We intend to start off with a database of meteorite recovery and work our way up the 'knowledge discovery in databases' framework to utilize data mining techniques and extract actionable insights from our data set. To define our project's scope, we have identified a set of research questions we would pursue.

The following is a list of research questions we have used to guide our research.

1) What value can be extracted from existing meteorite and geospatial data to aid in locating new meteorites?
2) What sort of biases and challenges exist in the data? How can we overcome these?
3) Are there any useful ways to visualize the data in a way that will help us identify patterns in the meteorite findings?

mds
September 24, 2023

1.1 Problem

Meteorites are a valuable resource for astronomical research as they provide insight into the composition of bodies in our solar system and beyond. The information can be used to find traces of water, potential resources for mining, and insights into the origin of our solar system. Being able to harvest meteorites more efficiently to gather more research data could prove beneficial.

This project aims to solve the problem of there not being an existing framework to find cost-effective locations to mine for meteorites. Factors that can impact this may include whether the area has already been explored for meteorites, proximity to accessible roads, and hazards such as mountains and bodies of water. There is a lot of data about meteorites and where they were located but the data is biased towards countries and regions that have had the most meteorite research done.

1.2 Data

We are going to use the meteorite landing dataset that contains columns such as Meteorite name, Unique ID, Name type (Valid / Relict), Recovery class, mass (g), Fall type (Fell / Found), Year, Recovery latitude location, Recovery Longitude location, and Geolocation coordinates. Shape 45717 rows and 10 columns. We will also overlay and combine public geospatial data.

1.3 Data Mining Task

Our objective is to highlight the sparseness in this dataset and explain the influence of degree of sparse dataset on the overall quality of the data mining effort. We will deploy a plethora of data preprocessing, exploratory data analysis, feature engineering, machine learning, and visualization to expose important patterns in the dataset. We will also combine this data with public geographical data.

The following is an organized list of tasks we aim to accomplish with our datasets.

1) We will filter data by Limiting the dataset to the geographical region we are interested in, such as the United States and Europe.
2) We will clean the data by removing any inconsistencies or outliers.
3) We will perform exploratory data analysis which includes:
    a) generating a heat map to visualize the density of meteorite findings across the region.
    b) Overlay various other geospatial data to try to identify patterns between meteorite findings and other geographical features.
4) Feature-engineering methods.
    a) We will use methods such as binning to discretize the geospatial data so that it can be more easily worked with.
    b) We will explore ways to combine meteorite data with public geospatial data to develop new features that might aid in our machine learning models.
5) Machine learning models to highlight searchable zones.

a) We will explore clustering models such as K-means and DBSCAN to identify high density regions of past meteorite findings and begin highlighting those as already explored regions.

b) We will explore the use of prediction models to predict the likelihood of finding a meteorite in unexplored locations based on features we identify during our exploratory data analysis phase.

6) Evaluation of our models.

a) We can employ means to test the efficacy of the models we developed by splitting the existing data into parts that can be used for validation.

b) We can run predictions and see how close that is to actual findings.

c) We can look at recent finds and see how likely our models estimate a find in that location to see if the model is generalizable to data outside of the dataset provided.

7) Post-processing and visualization.

a) The end goal of this data processing is to create an interpretable map of where to go digging for meteorites.

b) We plan to use either binning or other methods to visualize these results in a geospatial context.

1.4 Work Plan

(full gantt chart included with submission in separate PDF. Table available here in Fig. 1.)

2 Conclusion

Appendix A

Appendix B

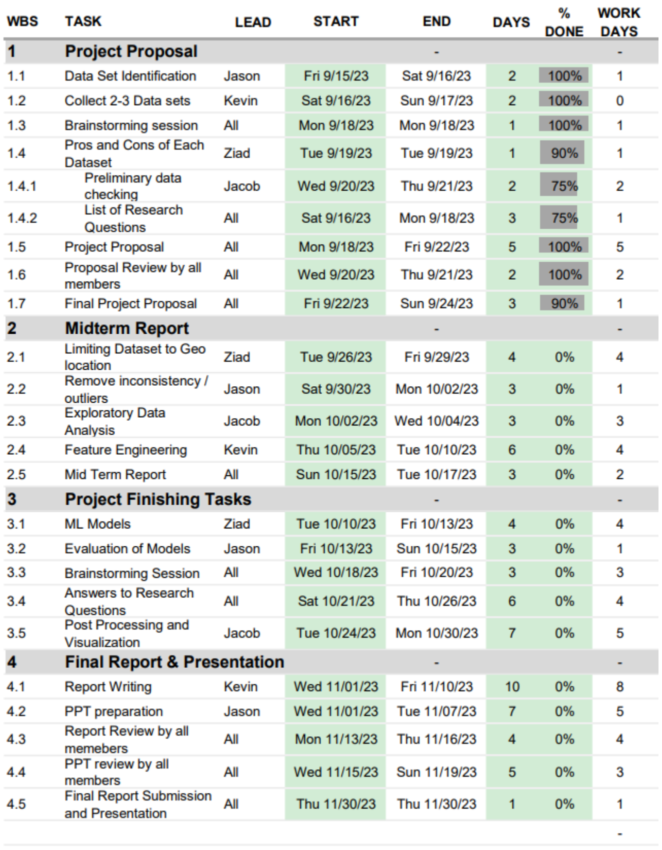
Acknowledgments

References

[1]

Ziad Arafat Ziad Arafat is a computer science major at NMSU seeking a career in Artificial Intelligence and embedded programming.

Kevin Dhanapal Kaverinathan (Kevin) Dhanapal MS PMP, is a Ph.D. student with the Department of Management at the New Mexico State University, Las Cruces, NM, U.S.A. He received his M.S. degree in petroleum engineering from the University of Oklahoma, Norman, OK, U.S.A. and a B.S. degree in mechanical engineering from P.S.G. College of Technology, Coimbatore, India. He is a reviewer for A.O.M. Annual and Regional conferences. His research interests include Industry 4.0, Operations & Supply chain management, supply chain entrepreneurship, space economy. He is a member of DSI and PMI.



WBS	TASK	LEAD	START	END	DAYS	% DONE	WORK DAYS
1	**Project Proposal**			-			-
1.1	Data Set Identification	Jason	Fri 9/15/23	Sat 9/16/23	2	100%	1
1.2	Collect 2-3 Data sets	Kevin	Sat 9/16/23	Sun 9/17/23	2	100%	0
1.3	Brainstorming session	All	Mon 9/18/23	Mon 9/18/23	1	100%	1
1.4	Pros and Cons of Each Dataset	Ziad	Tue 9/19/23	Tue 9/19/23	1	90%	1
1.4.1	Preliminary data checking	Jacob	Wed 9/20/23	Thu 9/21/23	2	75%	2
1.4.2	List of Research Questions	All	Sat 9/16/23	Mon 9/18/23	3	75%	1
1.5	Project Proposal	All	Mon 9/18/23	Fri 9/22/23	5	100%	5
1.6	Proposal Review by all members	All	Wed 9/20/23	Thu 9/21/23	2	100%	2
1.7	Final Project Proposal	All	Fri 9/22/23	Sun 9/24/23	3	90%	1
2	**Midterm Report**			-			-
2.1	Limiting Dataset to Geo location	Ziad	Tue 9/26/23	Fri 9/29/23	4	0%	4
2.2	Remove inconsistency / outliers	Jason	Sat 9/30/23	Mon 10/02/23	3	0%	1
2.3	Exploratory Data Analysis	Jacob	Mon 10/02/23	Wed 10/04/23	3	0%	3
2.4	Feature Engineering	Kevin	Thu 10/05/23	Tue 10/10/23	6	0%	4
2.5	Mid Term Report	All	Sun 10/15/23	Tue 10/17/23	3	0%	2
3	**Project Finishing Tasks**			-			-
3.1	ML Models	Ziad	Tue 10/10/23	Fri 10/13/23	4	0%	4
3.2	Evaluation of Models	Jason	Fri 10/13/23	Sun 10/15/23	3	0%	1
3.3	Brainstorming Session	All	Wed 10/18/23	Fri 10/20/23	3	0%	3
3.4	Answers to Research Questions	All	Sat 10/21/23	Thu 10/26/23	6	0%	4
3.5	Post Processing and Visualization	Jacob	Tue 10/24/23	Mon 10/30/23	7	0%	5
4	**Final Report & Presentation**			-			-
4.1	Report Writing	Kevin	Wed 11/01/23	Fri 11/10/23	10	0%	8
4.2	PPT preparation	Jason	Wed 11/01/23	Tue 11/07/23	7	0%	5
4.3	Report Review by all memebers	All	Mon 11/13/23	Thu 11/16/23	4	0%	4
4.4	PPT review by all members	All	Wed 11/15/23	Sun 11/19/23	5	0%	3
4.5	Final Report Submission and Presentation	All	Thu 11/30/23	Thu 11/30/23	1	0%	1
							-

Fig. 1. Table highlighting our work plan.

Jason Ivey Biography text here.

Jacob Yoder Biography text here.