

TAPS Grief Journey with R and AI



Mansi Shakalya
mshakaly@purdue.edu



Ronak Bhagia
rbhagia@purdue.edu



Christine Chiang
chian133@purdue.edu



Daniel Gallagher
gallag97@purdue.edu



Matthew A. Lanham
lanhamm@purdue.edu

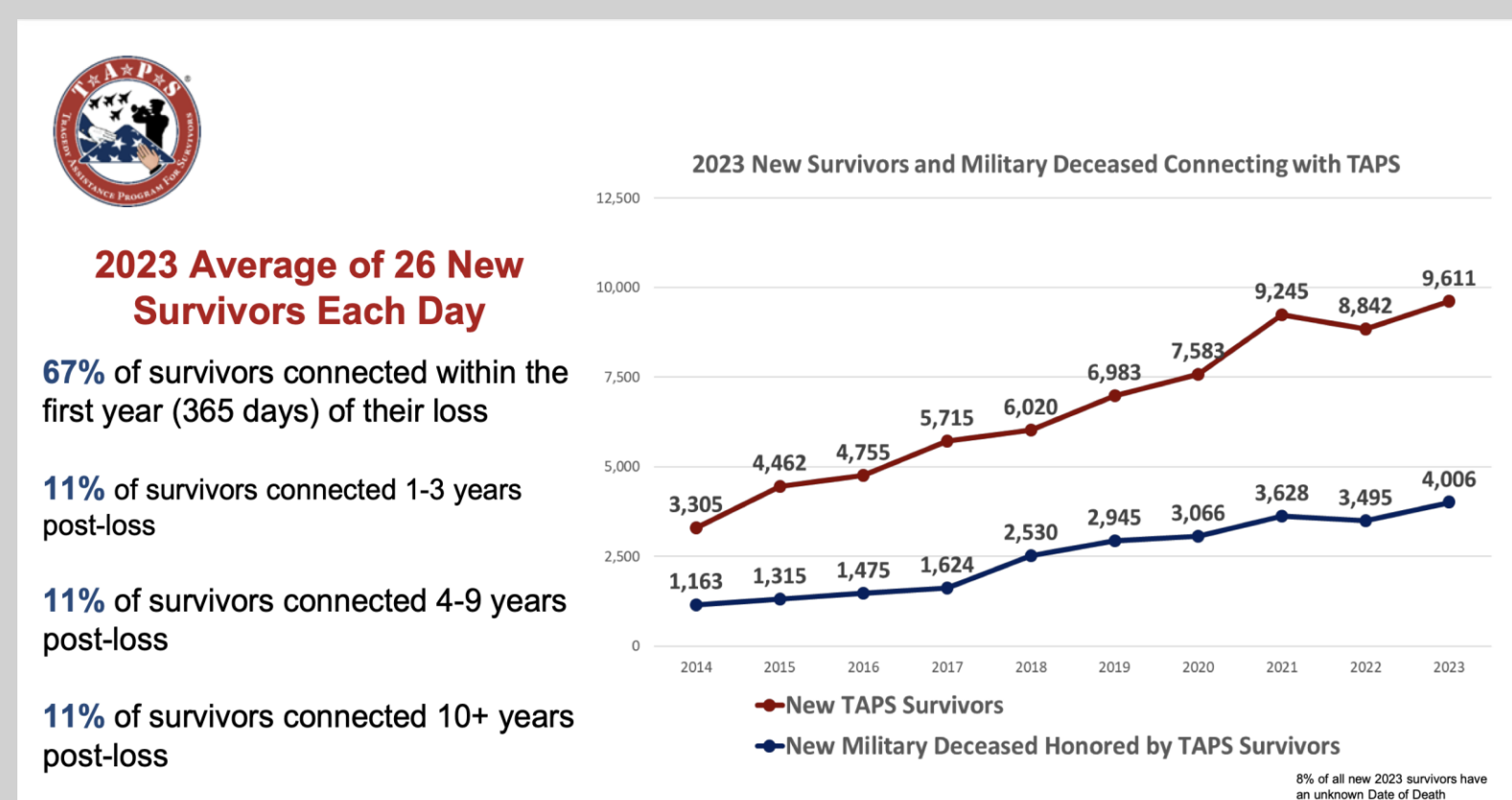
ABSTRACT

In our study we develop a data-driven solution for the Tragedy Assistance Program for Survivors (TAPS) that predicts the stage of grief that a person is in on a journey map designed by the people at TAPS. We predict the stages based on the answers someone gives on a survey they fill out after they attend a TAPS seminar using various data mining techniques and Microsoft Azure. The inspiration for this work comes from the idea of helping a person or military families by classifying them in the correct point in the journey and getting them the correct resources in order to help them improve their lives. Our solution involves applying cutting-edge AI and data science techniques to a real-world problem with significant social impact. This work research is part of the national Data4Good case competition.

BUSINESS PROBLEM

Since 1800, more than 37 million people worldwide have died while actively fighting in War (Our world in data, 2024). On average that is more than 170,000 people a year. The death of a loved one affects not only those closest to them but also extended family, friends etc. That is more than twice the people who are impacted due to a singular passing. Approximately 10% of beavered individuals will also develop a prolonged grief condition (Szuhany et al., 2021). Hence, the importance of recognizing the stage of grief a survivor is in, is integral to finding the appropriate grief counseling for their journey of healing. 87% of people attending TAPS seminars said it gave them hope for their future and identifying their stage of grief would help TAPS support them on their journey (TAPS Impact Report, 2023).

Sources:
Herre, B., Rod s-Guirao, L., & Roser, M. (2024, March 20). *War and Peace*. Our World in Data.
TAPS Impact Report 2023. (2024, March 19).
Szuhany, K. L., Malgaroli, M., Miron, C. D., & Simon, N. M. (2021). Prolonged Grief Disorder: Course, Diagnosis, Assessment, and Treatment. *Focus (American Psychiatric Publishing)*. 19(2), 161–172.

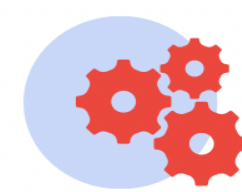


ANALYTICAL PROBLEM FRAMING



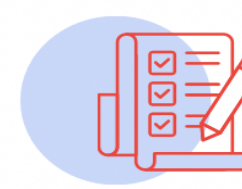
Problem Statement

The goal is to create a solution that will map the numeric and text responses from the survey to appropriate stages within the TAPS survivor journey map. The analytical problem is a **multi-classification** problem requiring us predict the categorical stage of grief.



Assumptions

Data is accurate and correct. Questions in the survey are relevant to accurately predict the stage of grief.



Success Metric

F-1 score: The harmonic mean of precision and recall, providing a balance between the two metrics. We are using the weighted F1 score to calculate metrics for each label, and find their average weighted by support (the number of true instances for each label).



Justification

The multi-classification approach was used since we are analyzing the response to identify the stage of grief of the individual. There is **one label per response** which is a stage of grief and they are **mutually exclusive** making this approach justified.

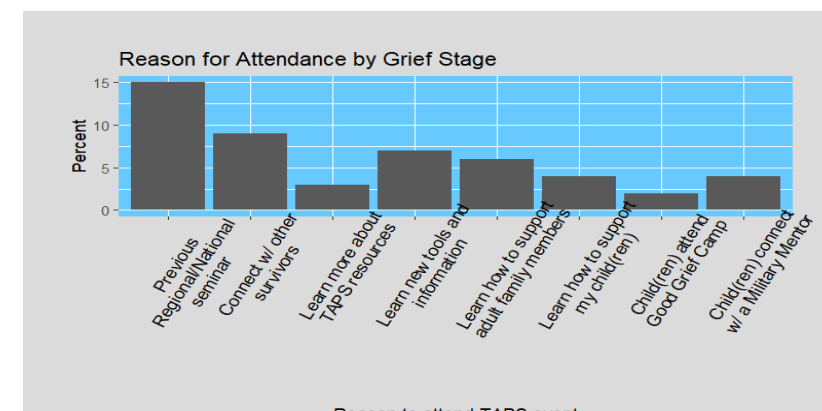
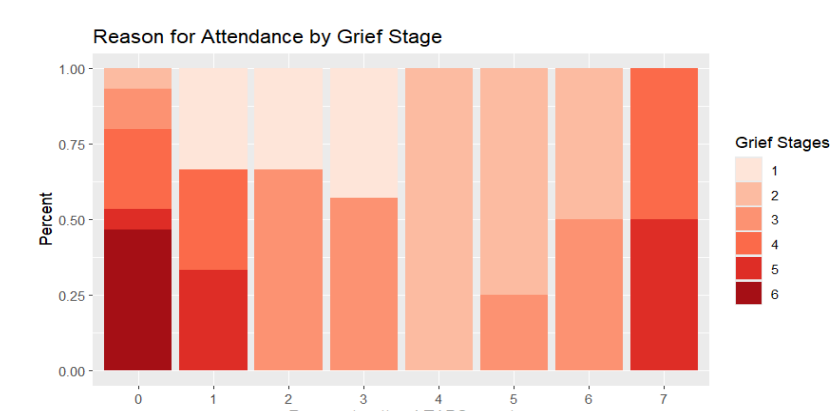
RESEARCH QUESTIONS

- What are the most effective methodologies for predicting an individual's specific stage of grief based on survey responses collected by the TAPS program
- Are there specific patterns or correlations within responses to particular survey questions that can be leveraged to more accurately categorize and predictions of an individual's stage of grief?

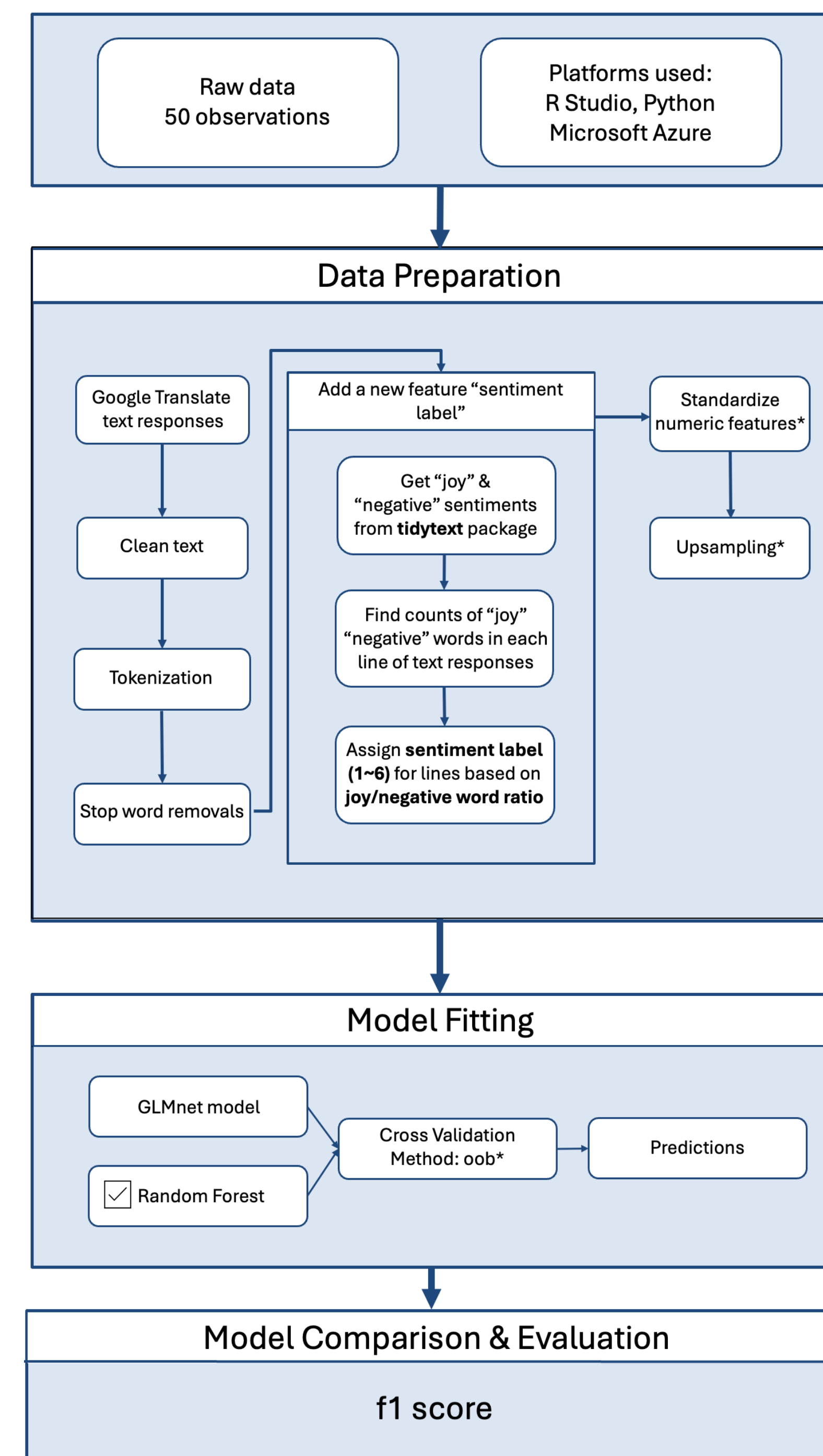
DATA

- Survey data collected through the TAPS program, with the goal of mapping responses to the corresponding stages of grief experienced by survivors.
- Labeled Data: A small dataset (tens of samples) where each response is associated with a specific stage in the grief journey, serving as a reference for model training.
- Unlabeled Data: A large dataset (thousands of responses) without predetermined labels, which will be used to classify individuals into their appropriate grief stage based on learned patterns.

GRAPHS



METHODOLOGY



***Standardize numeric features:** method "center" used to ensure features are on same scale to prevent bias. "Knninoute" is intergreted into the process to handle missing value

***Upsampling:** add more rows to the training dataset since there are only 50 observations, to train the modle better

***Cross validation, oob method:** Out-of-Bag data (OOB) - for each tree, the data points that were not selected in the bootstrap sample are referred to as "out-of-bag" data. These points can be used to evaluate the performance of the model without the need for a separate validation set

Random Forest yields a higher Accuracy and F1 Score

Cross Validation Technique: "LOOCV"
Classification Method: "GLMNET"

Accuracy	Kappa	Mean_F1
0.8939394	0.8727273	0.8863323
0.9090909	0.8909091	0.9046707
0.8787879	0.8545455	0.8724581
0.9090909	0.8909091	0.9053294
0.8939394	0.8727273	0.8862806
0.8636364	0.8363636	0.8518651
0.8939394	0.8727273	0.8738406
0.8787879	0.8545455	0.8603799
0.8484848	0.8181818	0.8387676

Cross Validation Technique: "oob"
Classification Method: "rf"

Accuracy	Kappa
0.9393939	0.9272727
0.9242424	0.9090909
0.8939394	0.8727273

0.93325

Business Implications

- Cost-Benefit: Having an accurate model can vastly improve the efficiency of getting the correct resources to the specific person and help save money by not giving the wrong resources and having to give up more time and money to give them the correct resource
- Azure: Using Microsoft Azure can help more accurately predict the stages of grief and would be a worthy investment to use this software to help improve models

Conclusions

- Accurately predicting the stage of grief is essential because being able to get the correct resources for each person in their grief journey can be critical to their progression
 - This can impact TAPS by providing them with a reliable model to continue to help millions of people around the world help better themselves while recovering from a loss of a loved one
- Limitations:
- Since the model had NA values these options didn't provide much to the model as it's difficult to say that an NA answer had any meaning to the final label so these could worsen the model's final prediction

Acknowledgements

We would like to thank Professor Matthew Lanham and TAPS for this opportunity, their guidance, and support on this project.