

10-25-2023

Cultivating the culture of responsible data science with Model-Cart

Vidit Singh

Yonas Kassa

Brian Ricks

Robin Gandhi

Follow this and additional works at: <https://digitalcommons.unomaha.edu/interdiscipinformaticsfacproc>

Please take our feedback survey at: [https://unomaha.az1.qualtrics.com/jfe/form/](https://unomaha.az1.qualtrics.com/jfe/form/SV_8cchtFmpDyGfBLE)

[SV_8cchtFmpDyGfBLE](https://unomaha.az1.qualtrics.com/jfe/form/SV_8cchtFmpDyGfBLE)

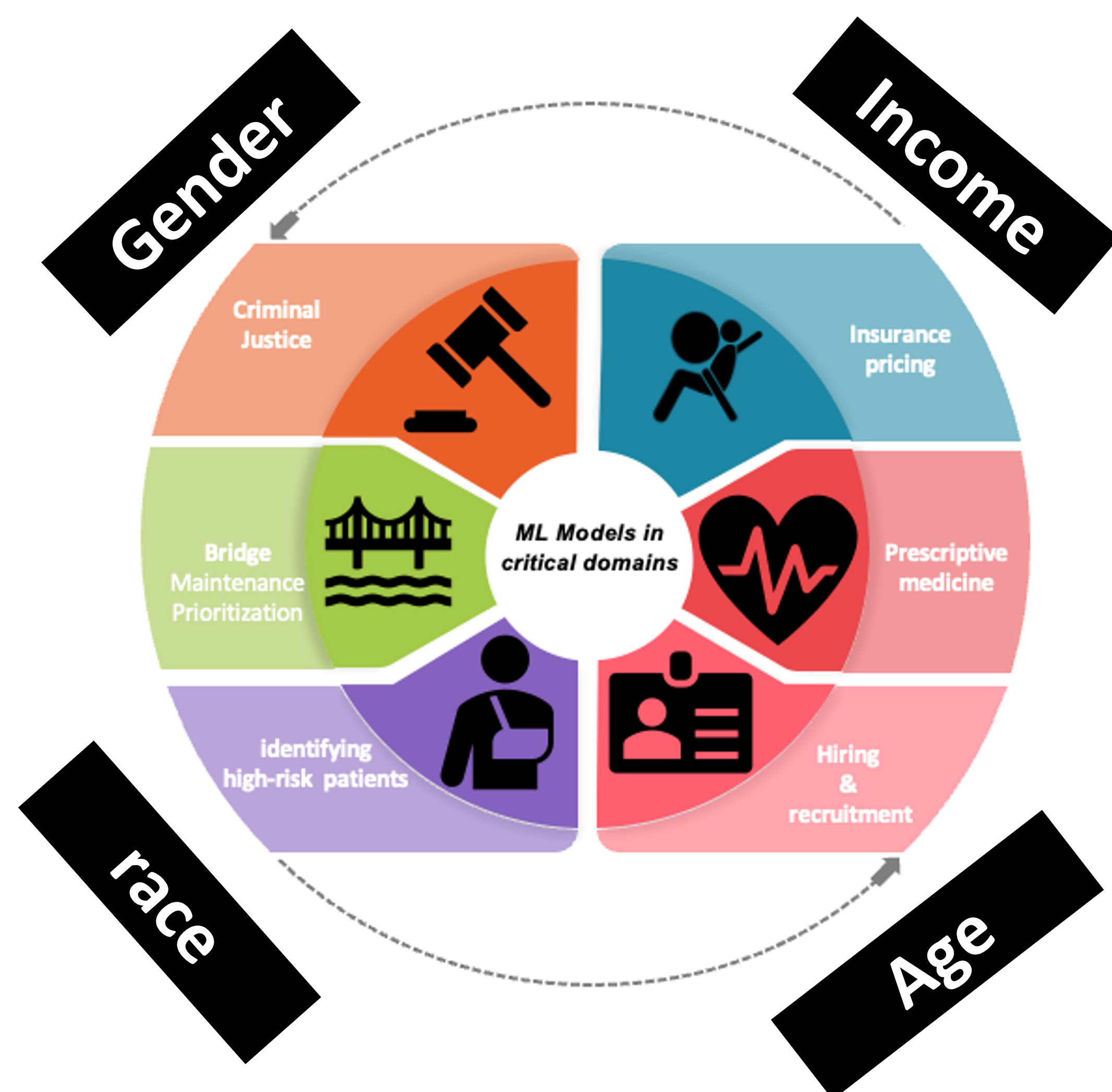


Cultivating the culture of responsible data science with Model-Cart: A Human-in-the-Loop approach to model training, evaluation, and deployment with Explainability

Vidit Singh, Dr. Yonas Kassa, Dr. Brian Ricks, Dr. Robin Gandhi

Machine Learning

Our work introduces *Model-Cart*, an explainable machine learning framework with human-in-the-loop that enables more reproducible and trustworthy data science. With a user-friendly interface and quantitative and qualitative model explainability techniques, our framework can improve the justifiability of ML model selection in high-stakes settings.



ML models are integral components of data science in multiple domains

Performance is just one dimension

Conclusion
The best performing machine learning culture results, and output proxies for provider:
which models perform better in predicting sandstorms.
The best performing model is used in our website.
results show that CART decision tree outperforms
...
n model with a rate of 0.820. The best performing model for
This approach could improve and ease the process of credit

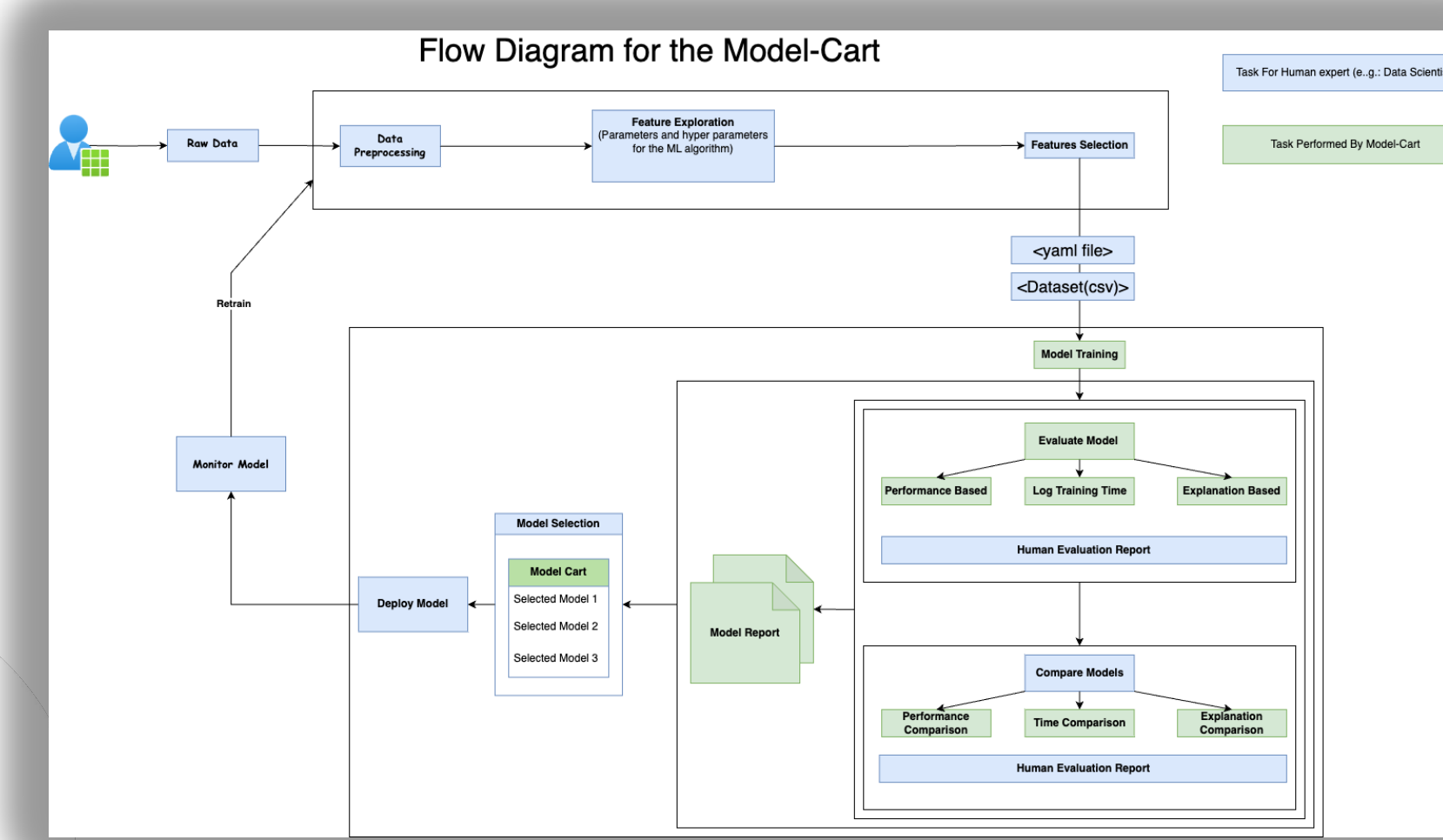


Model Training

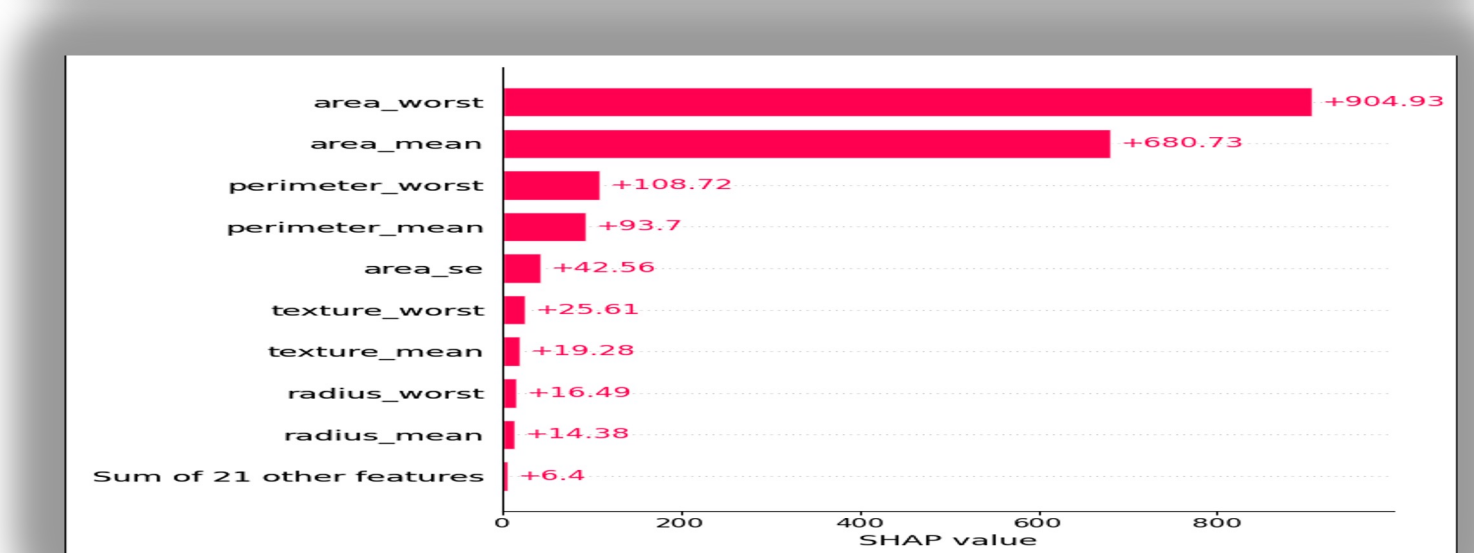
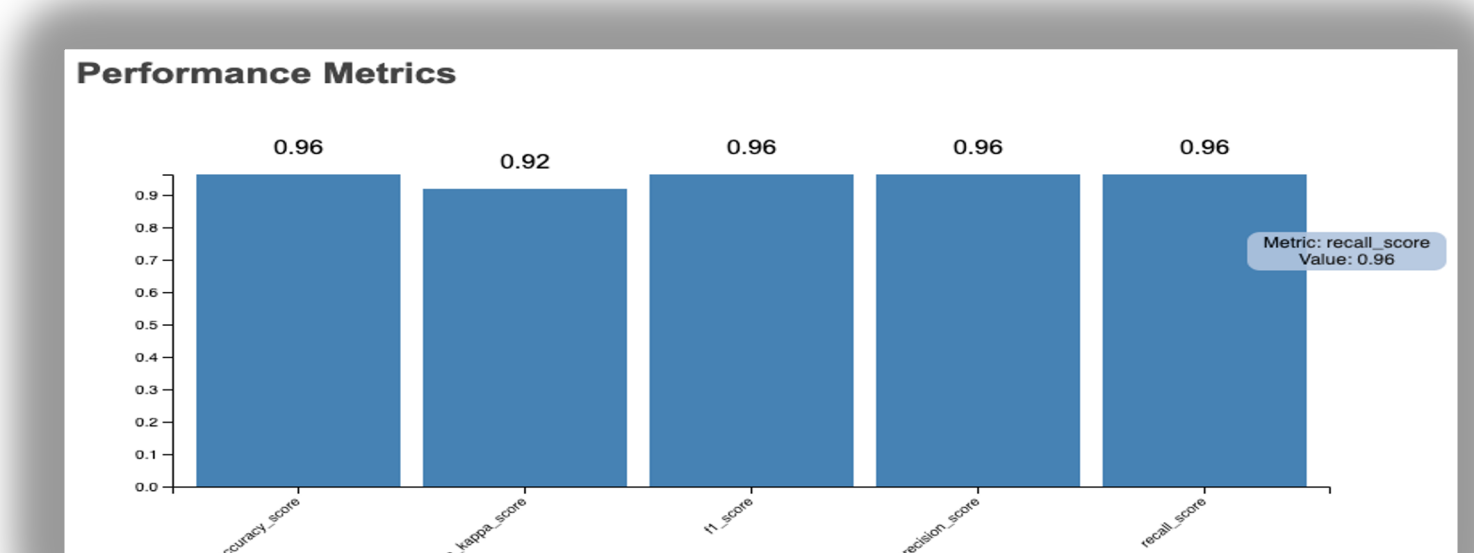
Built on *Miflow* to leverage state-of-the-art ML pipeline – streamlines training.

✓ Requires **dataset and model experiment specs only** – helps with reproducibility.

Trained ML models with quantitative evaluation results.



Model Evaluation: multiple dimensions considered



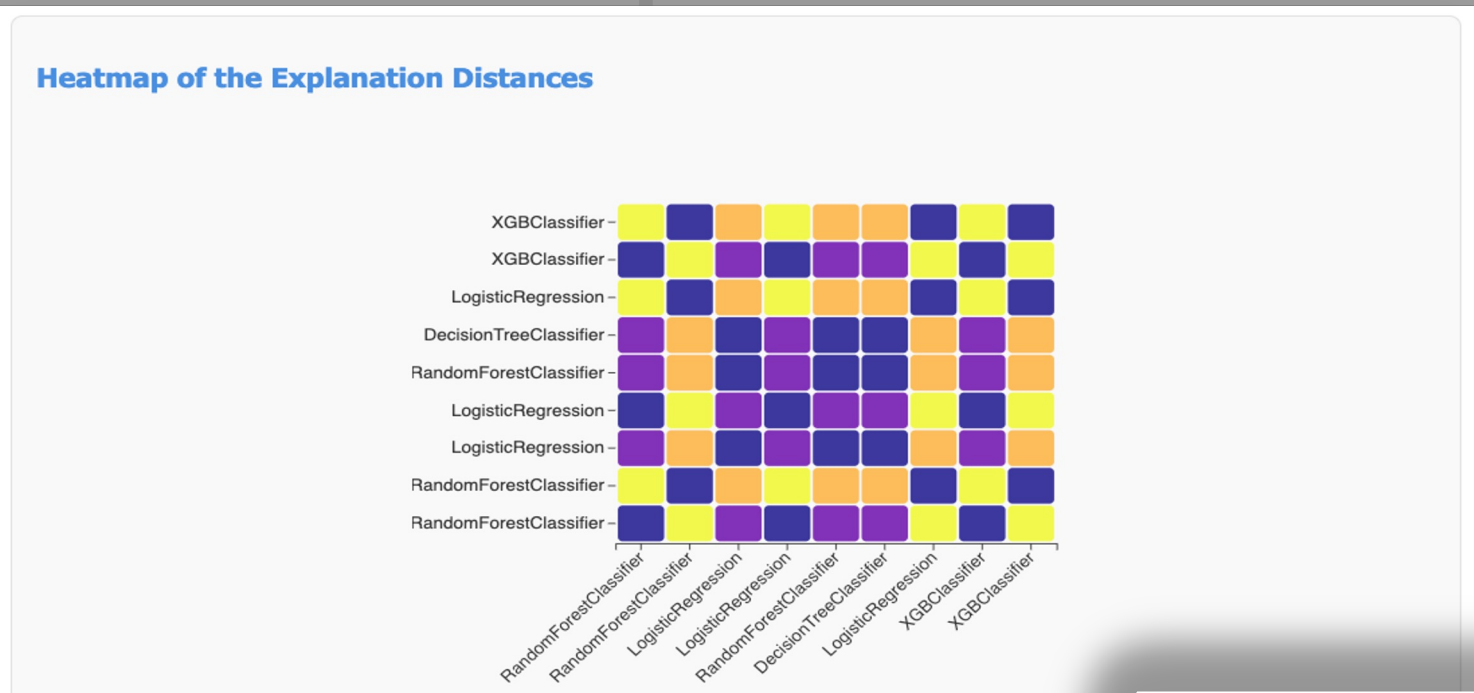
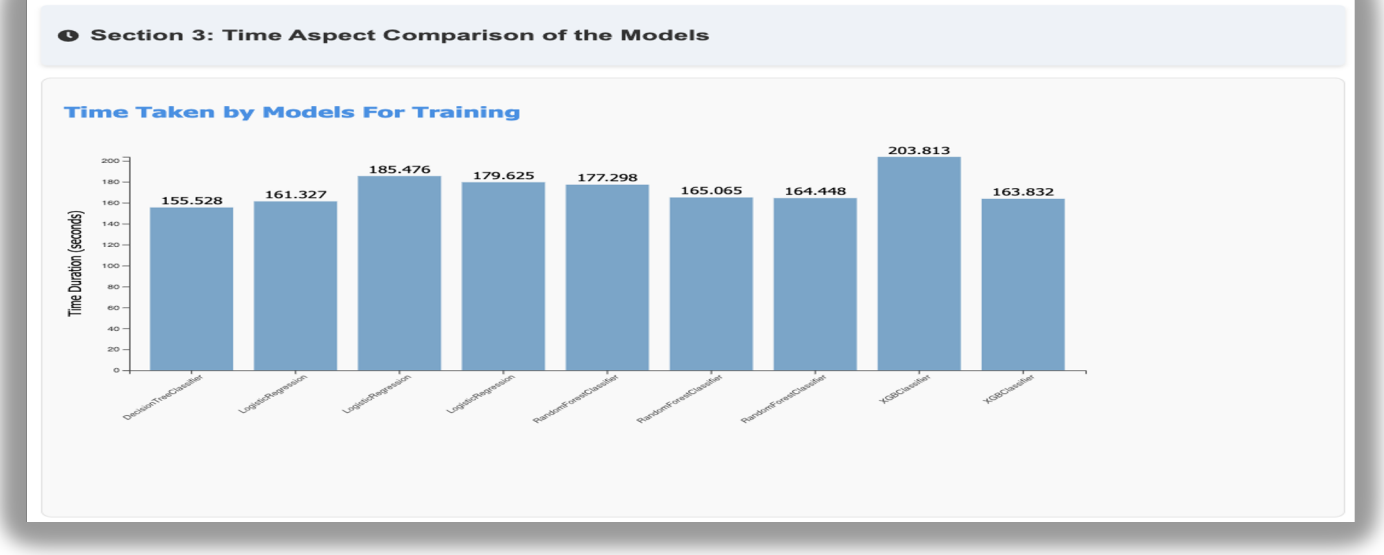
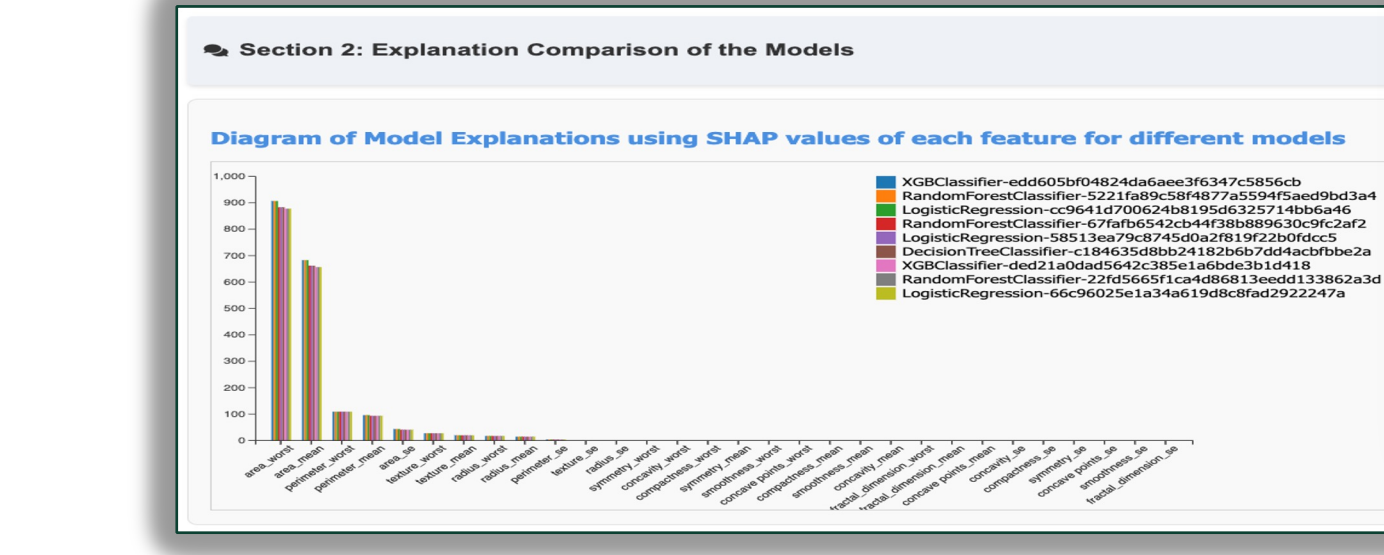
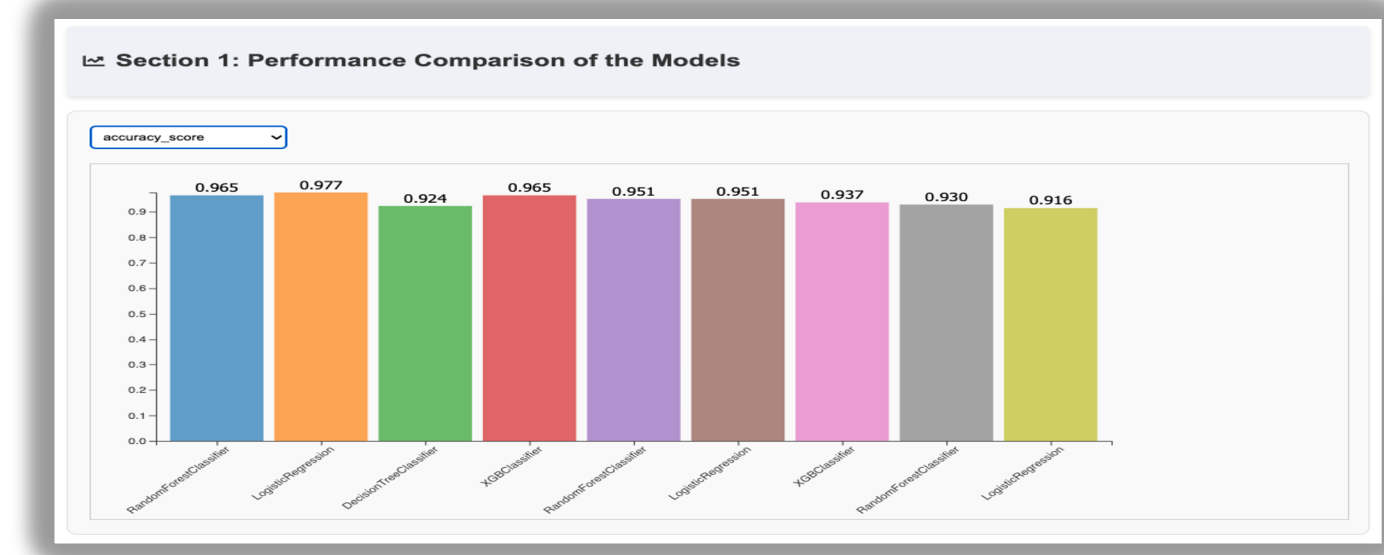
Human Expert Observation

ANALYST1 Notes

The XGBClassifier model seems to have been trained efficiently for a diagnostic task, achieving commendable accuracy. Its feature importance, as indicated by the SHAP values, points to certain parameters that are crucial for predictions. Depending on the context and application, some fine-tuning might further enhance its performance. Overall, the model appears well-optimized and robust.

Save Notes

Human-in-the-loop: Human decides which model and why?!



Section 4: Human Expert Observation

Comparison Notes

ANALYST1 Notes

Upon comparing the models, they all demonstrate comparable performance scores, indicating their similar proficiency on the dataset. While their SHAP values reveal variances in feature influence, the time taken for training remains in a close range. This assessment suggests that, depending on specific use cases or constraints (like real-time applications), one might choose a model with faster training times, or if interpretability is a priority, models with higher SHAP values for specific features might be preferred.

Save Comparison

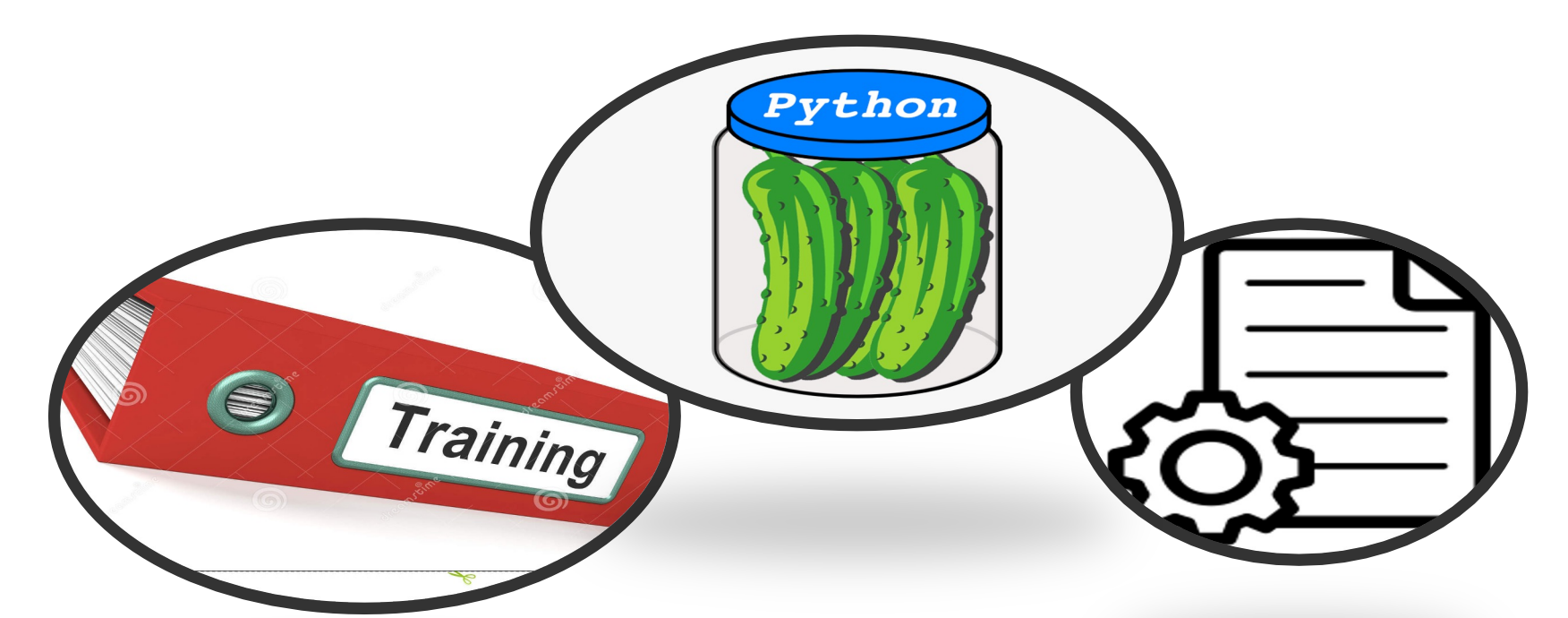
Model Deployment

Ready to deploy prioritized models with detailed model reports including a dataset report (w **pandas Profiling**).

Model Prioritization

Experiment ID	Run ID	Run Name	Priority	Action
99853735041708499	f452370a8b24c3d29f922a7d52f462	DecisionTreeClassifier	1	[icon]
99853735041708499	5d3935037ca4674b5e418c08a9795a8	RandomForestClassifier	2	[icon]
99853735041708499	e8976de43b24621b5a0a09e74e07abfd	LogisticRegression	3	[icon]

Add Selected to Cart



ANALYST1 - Cart

Experiment ID: 512852785488287218

ANALYST1 Experiment Comparison Notes

Upon comparing the models, they all demonstrate comparable performance scores, indicating their similar proficiency on the dataset. While their SHAP values reveal variances in feature influence, the time taken for training remains in a close range. This assessment suggests that, depending on specific use cases or constraints (like real-time applications), one might choose a model with faster training times, or if interpretability is a priority, models with higher SHAP values for specific features might be preferred.

Run ID: fu054e7688ef46e7979322a4f98ace52b Priority: 1 REMOVE DEPLOY

ANALYST1 Notes

The XGBClassifier model seems to have been trained efficiently for a diagnostic task, achieving commendable accuracy. Its feature importance, as indicated by the SHAP values, points to certain parameters that are crucial for predictions. Depending on the context and application, some fine-tuning might further enhance its performance. Overall, the model appears well-optimized and robust.

CLEAR CART

Acknowledgements

This work is partially supported by contracts W912HZ21C0060 and W912HZ23C0005, US Army Engineering Research and Development Center (ERDC), and Award Number 1762034 from the National Science Foundation.

