UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Interactive Visualization and Validation Techniques for Algorithmic Decision Making: Outcome-Explorer and SliceTeller Approaches

Rishabh Tiwari

Md Naimul Hoque and Klaus Mueller, Senior Member, IEEE Outcome-Explorer: A Causality Guided Interactive Visual Interface for Interpretable Algorithmic Decision Making

Xiaoyu Zhang SliceTeller : A Data Slice-Driven Approach for Machine Learning Model Validation, Jorge Piazentin Ono, Huan Song, Liang Gou, Kwan-Liu Ma, Liu Ren - VOL. 29, NO. 1, JANUARY 2023

Heidelberg, Germany, 1st March, 2023

VISUAL
COMPUTING
GROUP
HEIDELBERG UNIVERSITY

Introduction

- Supporting non-expert users in Explainable AI is very challenging.

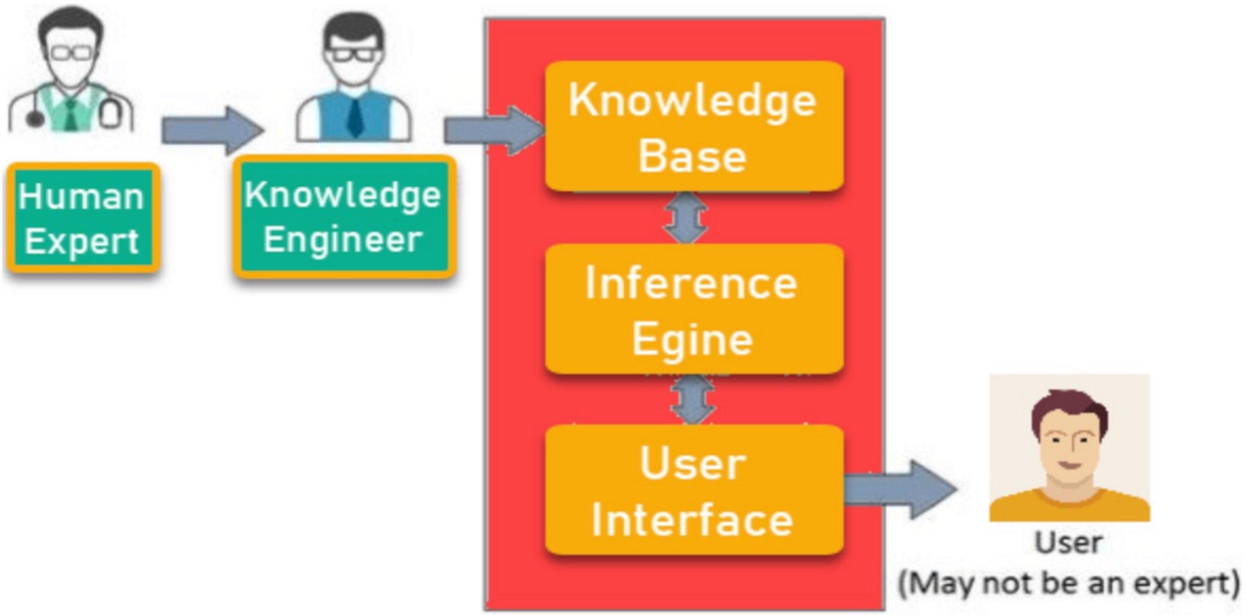
  - Different goals, reasons and skill sets for interpreting a ML model than expert users.

- A ML practitioner will test accuracy and fairness of the model
- A non-expert user wants to understand the functioning of the model to trust the results
- Machine Learning (ML) has a variety of critical applications.

  - Autonomous driving, medical imaging, industrial fire detection, etc.

- Applications need to be thoroughly evaluated before deployment

  - It may cause serious consequences

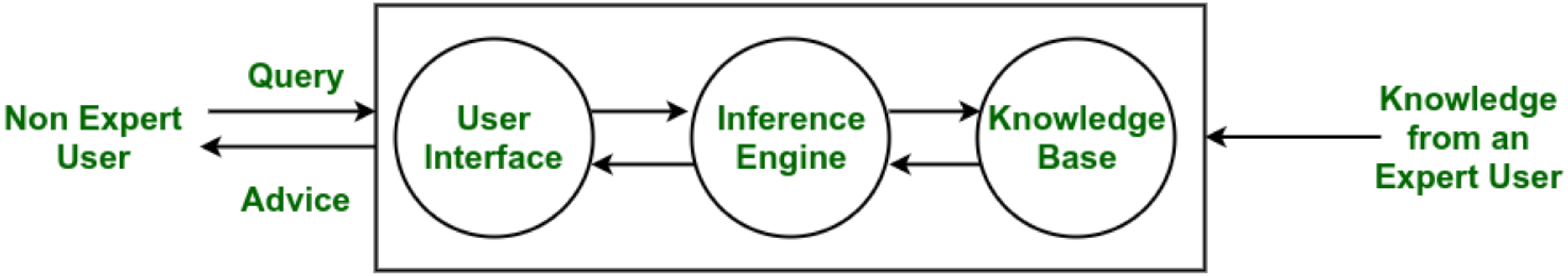


[1]

[1] https://www.guru99.com/expert-systems-with-applications.html

Motivation

- ML models, can be complex and difficult for users to understand and interpret
- Transparent decision-making is critical, such as in fields like medicine finance, and law.
- Outcome-Explorer solves the problem of interpretability and transparency in decision-making algorithms
- Real-world ML applications need to be thoroughly evaluated for model release
  - To ensure fairness and to achieve a consistent performance in various scenarios
- Discover where models fail, understanding why they fail and mitigating these problems
- SliceTeller, a tool that allows users to debug, compare and improve ML models driven by critical data slices.



[1]

[1] https://www.geeksforgeeks.org/expert-systems/

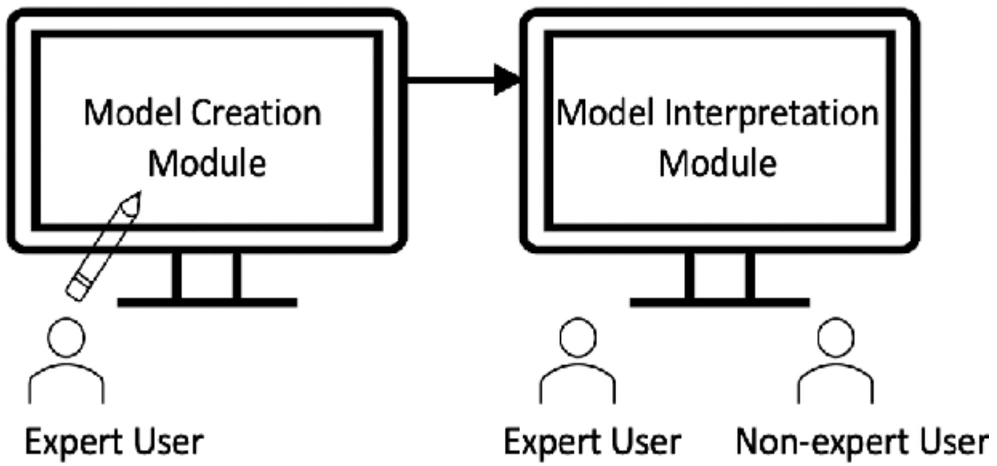
Problem Setting – Outcome Explorer

- Traditional algorithms can be difficult for even experts to comprehend due to their opaque nature.

- It is challenging to understand how input factors influence the algorithm's output and decision-making process.

- Outcome-Explorer addresses this problem by providing a visual interface and causality-guided interactions

- Allows users to explore and understand the decision-making
- Makes the algorithm's decision-making process more interpretable
- Provides the user with a visual representation of the process
- Allows the user to explore different outcomes
- Helps the user to gain deeper understanding of how the algorithm works



[1]

1. IEEE TRANSACTION ON VISUALIZATION AND COMPUTER GRAPHICS, 2021. AUTHOR'S VERSION. HTTPS://DOI.ORG/10.1109/TVCG.2021.3102051

Problem Setting - SliceTeller

- Manually slicing the data is a very time-consuming task.

  - Experts mentioned that this task involved manually creating rules to slice the data

  - Running evaluation scripts on the data subsets and comparing the results on various data subsets.

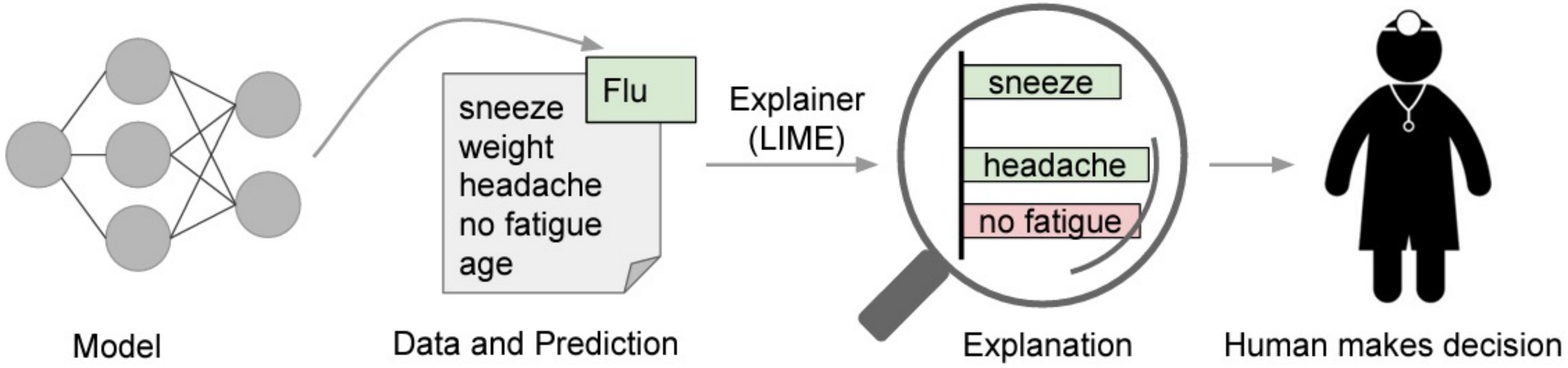
- ML experts cannot explore all possible data subsets to identify relevant failure cases for their application.

- Data slices can be created by any number of interpretable meta-data (e.g., weather and temperature for autonomous driving), resulting in an exponentially large search space.

Related Work
Method I
Outcome Explorer

Related Work

LIME (Local Interpretable Model-Agnostic Explanations) method:



[1]

Aim

- Addresses "black box" nature of many ML models.
- It is difficult to understand why a particular prediction was made.

Drawback

- It will be very difficult for a non-expert user to understand its function and gain trust in this model.

[1] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should i trust you? explaining the predictions of any classifier,” in Proc. ACM Knowledge Discovery and Data Mining, 2016, pp. 1135 1144

Related Work

- The approach makes causal analysis more accessible and intuitive for non-expert users.

- Provides them with visual representations of the data and relationships.

- Improves the ability of non-expert users to make data-driven decisions.

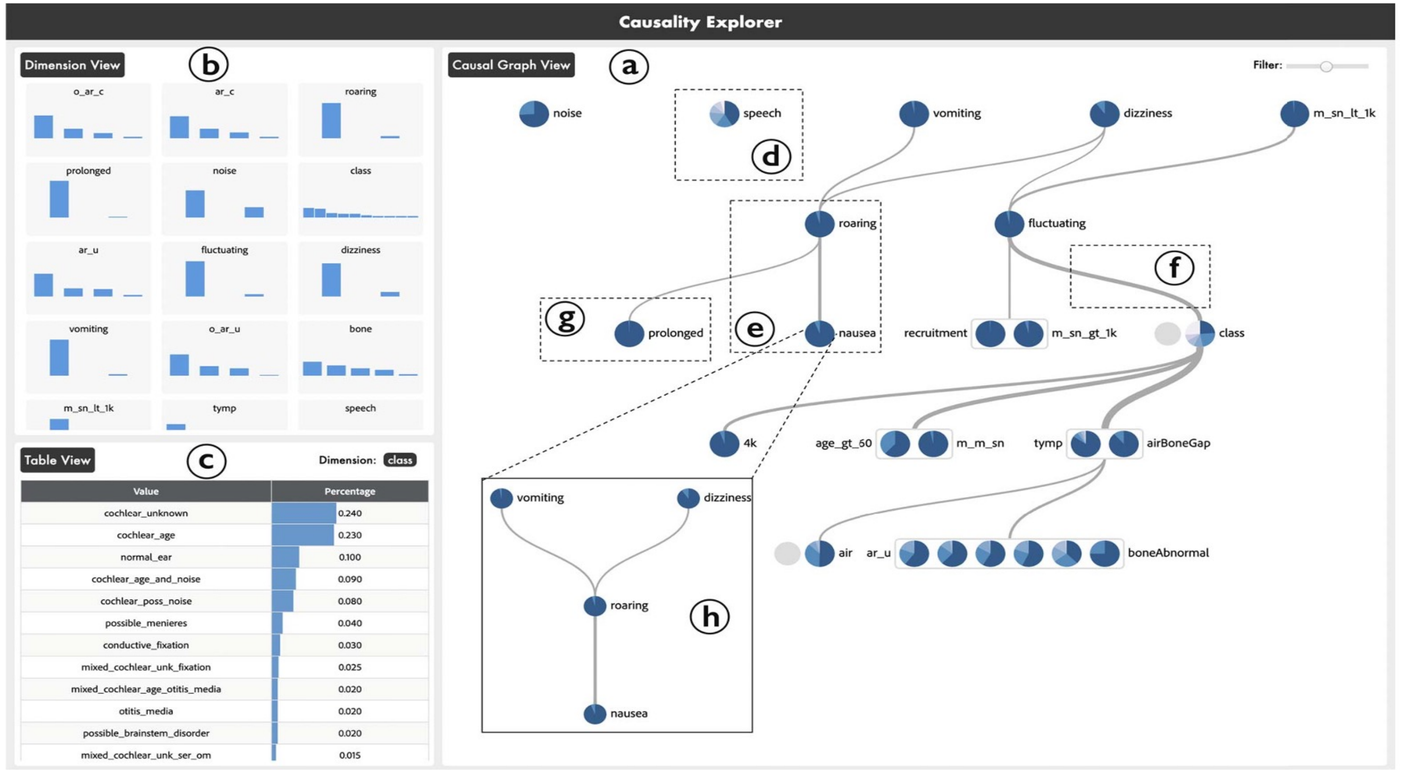


Table View c Dimension: class

Value	Percentage
cochlear_unknown	0.240
cochlear_age	0.230
normal_ear	0.100
cochlear_age_and_noise	0.090
cochlear_poss_noise	0.080
possible_menieres	0.040
conductive_fixation	0.030
mixed_cochlear_unk_fixation	0.025
mixed_cochlear_age_otitis_media	0.020
otitis_media	0.020
possible_brainstem_disorder	0.020
mixed_cochlear_unk_ser_om	0.015

[1]

Limitation

- It neglects temporal variables. Variables in these cases are all static.

- Integration of users' domain knowledge. Users still want interactive causal graph customization.

Outcome Explorer is more amenable to non-expert users who do not think in terms of distributions, uncertainties, and probabilities.

[1] Xiao Xie, Fan Du, and Yingcai Wu, " A visual analytics approach for exploratory causal analysis: Exploration, validation and applications"

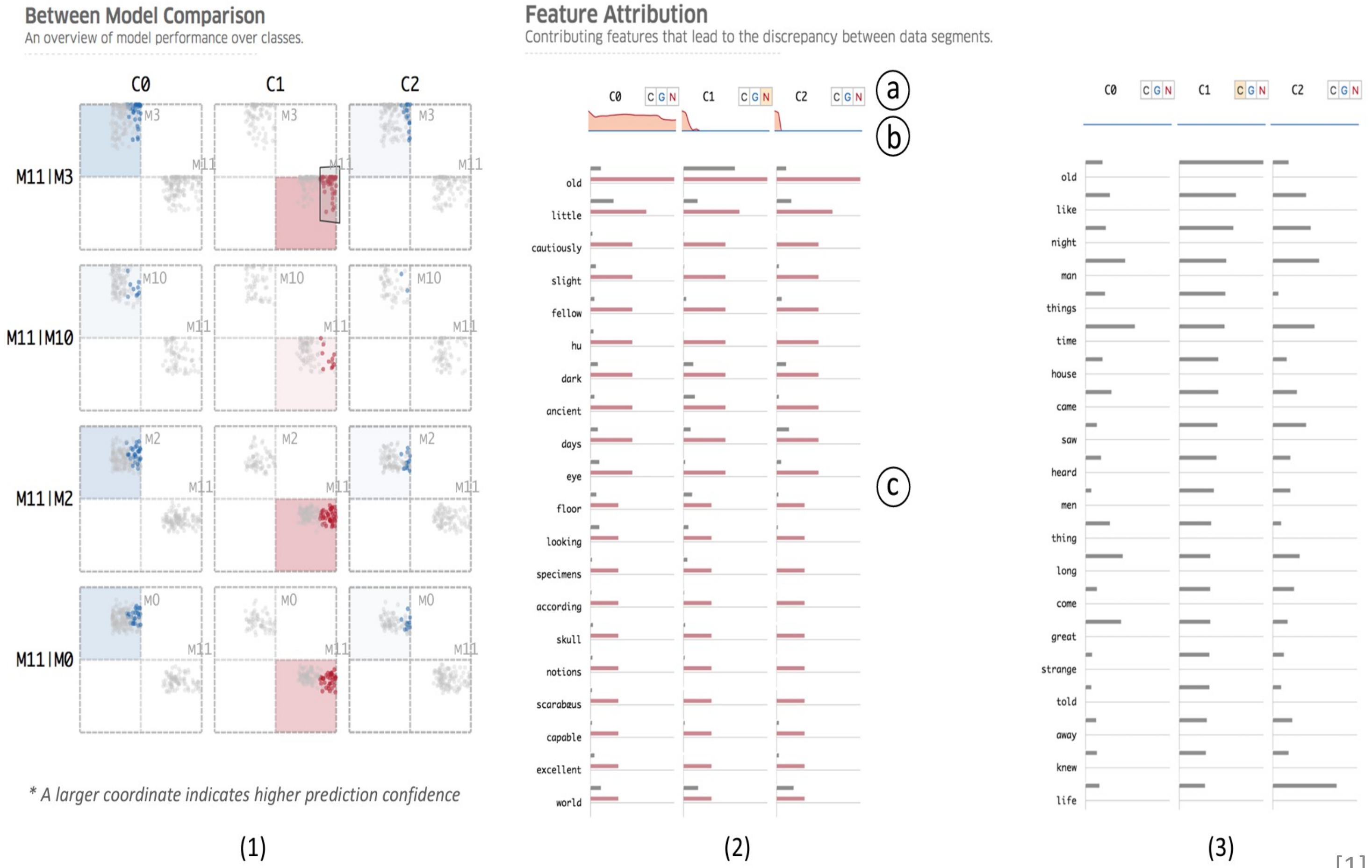
Related Work
Method II
(SliceTeller)

Related Work

- Manifold represents the relationship between the input data and the model's prediction as a low-dimensional manifold.
- Allows users to understand the model's behaviour and diagnose its performance in various scenarios.

Aim:

- To provide a more intuitive and visual way of understanding.
- Allowing users to explore.
- Diagnose the model's performance in real-time.
- It is designed to be flexible and easy to use.
- Makes it accessible to a wide range of users.
- Also including non-experts.

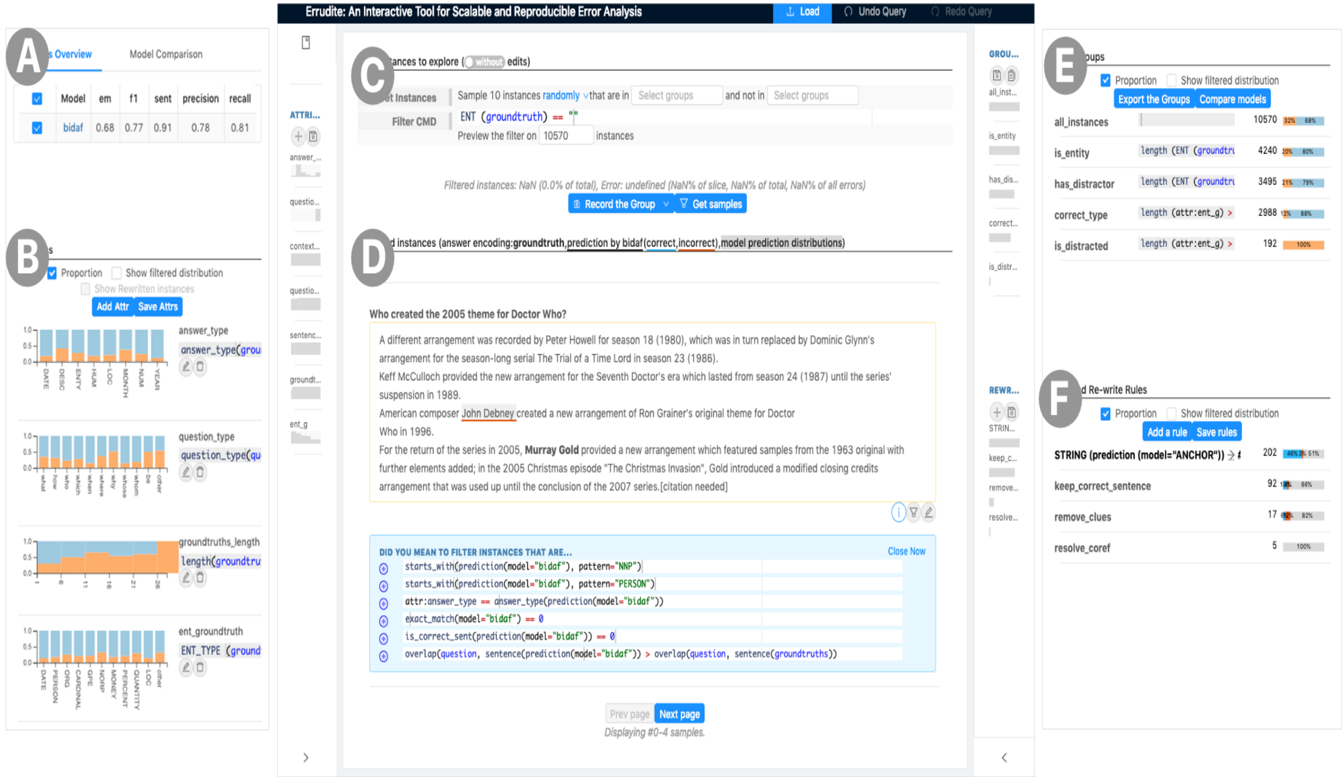


[1]

[1] 1. Manifold: A Model-Agnostic Framework for Interpretation and Diagnosis of Machine Learning Models Jiawei Zhang, Yang Wang, Piero Molino, Lezhi Li and David S. Ebert, Fellow, IEEE

Related Work

- It enables users to perform high quality
- Reproducible error analyses with less effort.
- Enhances the error analysis experience
- Allowing users to test and revise prior beliefs.



☑	Model	em	f1	sent	precision	recall
☑	bidaf	0.68	0.77	0.91	0.78	0.81

```
answer_type(grou
question_type(qu
length(groundtru
ENT_TYPE (ground
```

ances to explore (without edits)

et Instances: Sample 10 instances randomly that are in Select groups and not in Select groups

Filter CMD: `ENT (groundtruth) == ""`

Preview the filter on 10570 instances

Filtered instances: NaN (0.0% of total), Error: undefined (NaN% of slice, NaN% of total, NaN% of all errors)

d instances (answer encoding:groundtruth,prediction by bidaf(correct,incorrect),model prediction distributions)

Who created the 2005 theme for Doctor Who?

A different arrangement was recorded by Peter Howell for season 18 (1980), which was in turn replaced by Dominic Glynn's arrangement for the season-long serial The Trial of a Time Lord in season 23 (1986).
Keff McCulloch provided the new arrangement for the Seventh Doctor's era which lasted from season 24 (1987) until the series' suspension in 1989.
American composer John Debney created a new arrangement of Ron Grainer's original theme for Doctor Who in 1996.
For the return of the series in 2005, **Murray Gold** provided a new arrangement which featured samples from the 1963 original with further elements added; in the 2005 Christmas episode "The Christmas Invasion", Gold introduced a modified closing credits arrangement that was used up until the conclusion of the 2007 series.[citation needed]

DID YOU MEAN TO FILTER INSTANCES THAT ARE...

```
starts_with(prediction(model="bidaf"), pattern="NNP")
starts_with(prediction(model="bidaf"), pattern="PERSON")
attr:answer_type == answer_type(prediction(model="bidaf"))
exact_match(model="bidaf") == 0
is_correct_sent(prediction(model="bidaf")) == 0
overlap(question, sentence(prediction(model="bidaf")) > overlap(question, sentence(groundtruths))
```

Displaying #0-4 samples.

all_instances		10570	32% 68%
is_entity	`length (ENT (groundtru`	4240	20% 80%
has_distractor	`length (ENT (groundtru`	3495	21% 79%
correct_type	`length (attr:ent_g) >`	2988	12% 88%
is_distracted	`length (attr:ent_g) >`	192	100%

STRING (prediction (model="ANCHOR")) → #	202	46% 3% 51%
keep_correct_sentence	92	10% 4% 86%
remove_clues	17	6% 12% 82%
resolve_coref	5	100%

[1]

Limitation :

- Main focus: search efficiency and scalability
- Ignore how to understand and interpret the impact of these data slices.
- Errudite provides a domain-specific language for data grouping with unstructured text data analysis

1. Errudite: Scalable, Reproducible, and Testable Error Analysis Tongshuang Wu, Marco Tulio Ribeiro, Jeffrey Heer, and Daniel S. Weld

Fundamentals

- Visual Interface - They make extensive use of graphical objects.

- A user may directly manipulate the screen through several kinds of pointing devices (including her/his fingers).

- Get an almost instantaneous feedback (near real-time interactivity).

- Interactive Interface - Allows users to interact with a computer system or software in a dynamic and responsive way.

- Accepts user input, allowing the user to control the system.

- Provides dynamic feedback to the user, responding to user input.

[1]

[1] https://www.vecteezy.com/vector-art/2741565-man-with-store-terminal-flat-color-vector-faceless-character-contactless-payment-machine-isolated-cartoon-illustration-on-white-background-supermarket-self-service-kiosk-with-interactive-interface

Fundamentals

- Causal Guided Interactive Interface - Users interact with the system guided by principles of causality.

- Relationship between an event (the cause) and a second event (the effect), where the second event is a result of the first.

- Easier for users to understand and predict the effects of their actions.

- Data Slice Driven Approach - Analysing and processing data where data is divided into smaller and manageable slices.

- Each slice is analysed and processed separately.

- ML Model Validation - Process of evaluating the performance of a ML model on a set of data that it has not seen during training.

- Estimates the performance of the model on unseen data.

Method I - Background

- Pearl’s Structural Causal Model (SCM) expressed as DAG - Directed acyclic graphs (path diagrams).
- Variables are categorised as either exogenous (U) (Independent with no parents) or endogenous (v) – causal effects of exogenous variables.
- A,B - exogenous; C,D,E - endogenous;

$$E = \beta_{CE}C + \beta_{DE}D$$

$$E = \beta_{CE}C + \beta_{DE}(\beta_{AD}A + \beta_{BD}B)$$

$$E = \beta_{CE}(\beta_{AC}A + \beta_{BC}B) + \beta_{DE}(\beta_{AD}A + \beta_{BD}B)$$

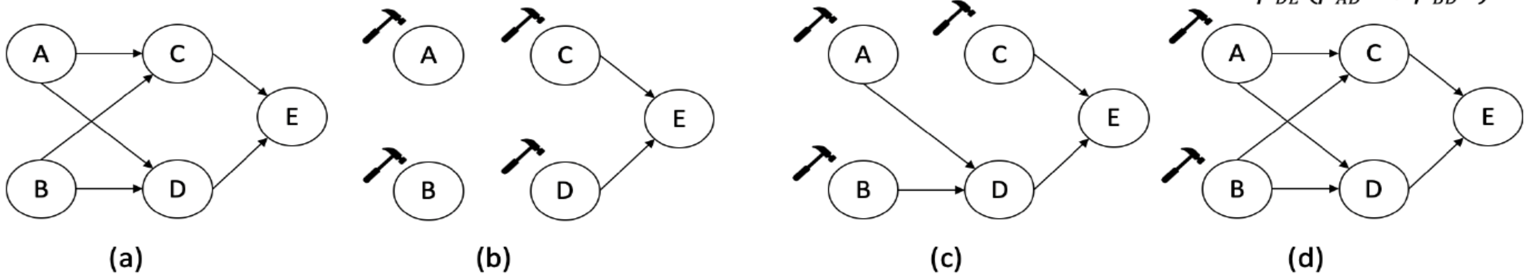


Fig. 1. Prediction in a causal model. The hammer icon represents intervention. (a) true causal model. (b) Interventions on all feature variables. The causal links leading to node C and D are removed since the values of C and D are set externally. (c) Interventions on node A, B, and C. (d) interventions on node A and B. In the path equations above models (b)-(d), the β are standardized regression coefficients estimated from the data.

[1]

[1] IEEE TRANSACTION ON VISUALIZATION AND COMPUTER GRAPHICS, 2021. AUTHOR’S VERSION. HTTPS://DOI.ORG/10.1109/TVCG.2021.3102051

Method I – Formative study with non expert users:

- Need for Transparency - users prefer an automated system over human assistance .

- Automated systems provide results efficiently and quickly.

- Human assistance takes a lot of time for the same.

- Improving Decision - Need of human assistance even after the automated results.

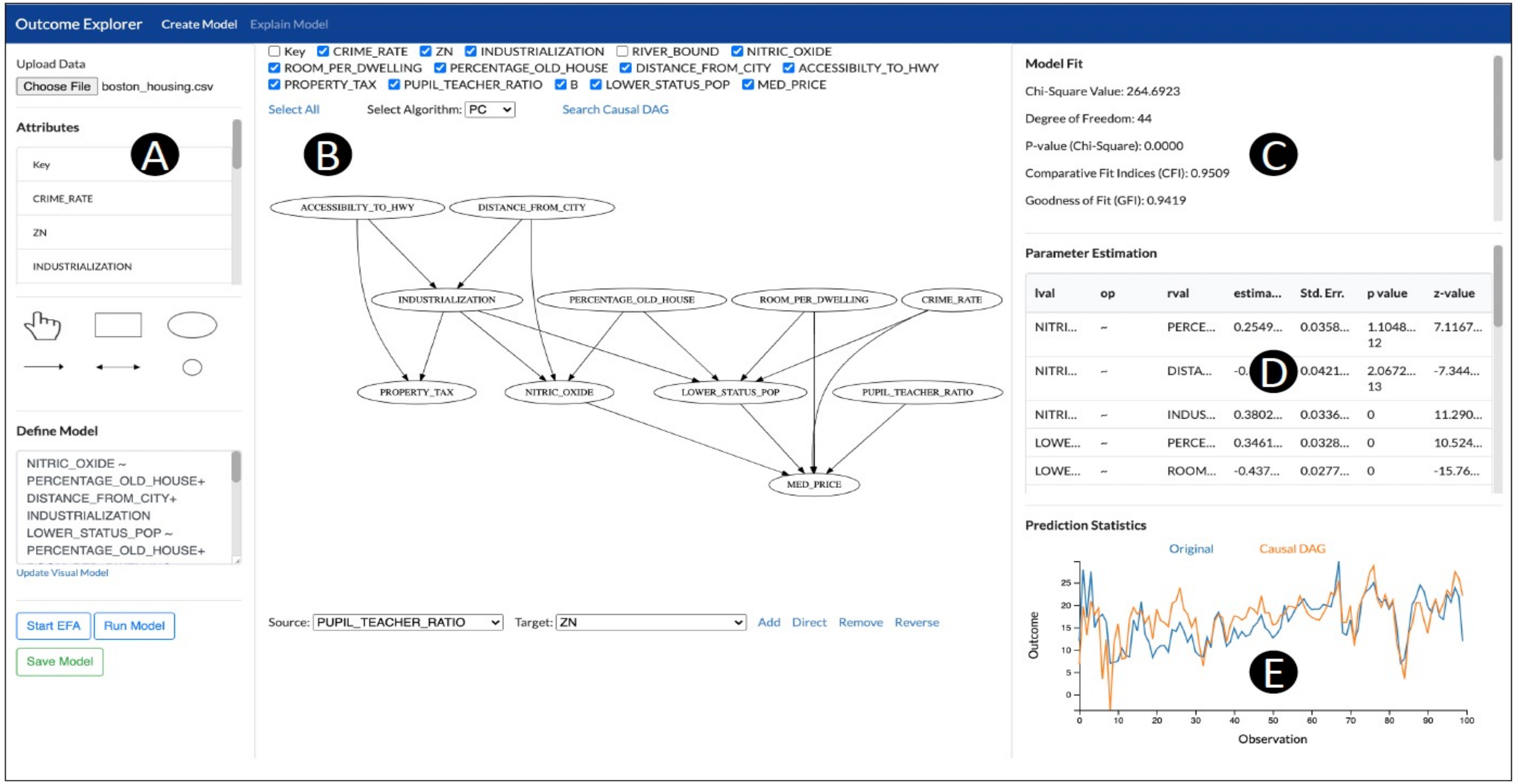
- Rules and specifications were not easily readable.

- Automated systems are complex for a non-expert user to understand.

- Different Decisions - participants received a decision different from their friend’s.

- Reason: Capabilities of one user is always different from the other.

Method I – Module Design



Upload Data
Choose File boston_housing.csv

□ Key ☑ CRIME_RATE ☑ ZN ☑ INDUSTRIALIZATION □ RIVER_BOUND ☑ NITRIC_OXIDE
☑ ROOM_PER_DWELLING ☑ PERCENTAGE_OLD_HOUSE ☑ DISTANCE_FROM_CITY ☑ ACCESSIBILTY_TO_HWY
☑ PROPERTY_TAX ☑ PUPIL_TEACHER_RATIO ☑ B ☑ LOWER_STATUS_POP ☑ MED_PRICE

Select Algorithm: PC

Define Model

NITRIC_OXIDE ~
PERCENTAGE_OLD_HOUSE+
DISTANCE_FROM_CITY+
INDUSTRIALIZATION
LOWER_STATUS_POP ~
PERCENTAGE_OLD_HOUSE+

Source: PUPIL_TEACHER_RATIO Target: ZN

Model Fit

Chi-Square Value: 264.6923

Degree of Freedom: 44

P-value (Chi-Square): 0.0000

Comparative Fit Indices (CFI): 0.9509

Goodness of Fit (GFI): 0.9419

Parameter Estimation

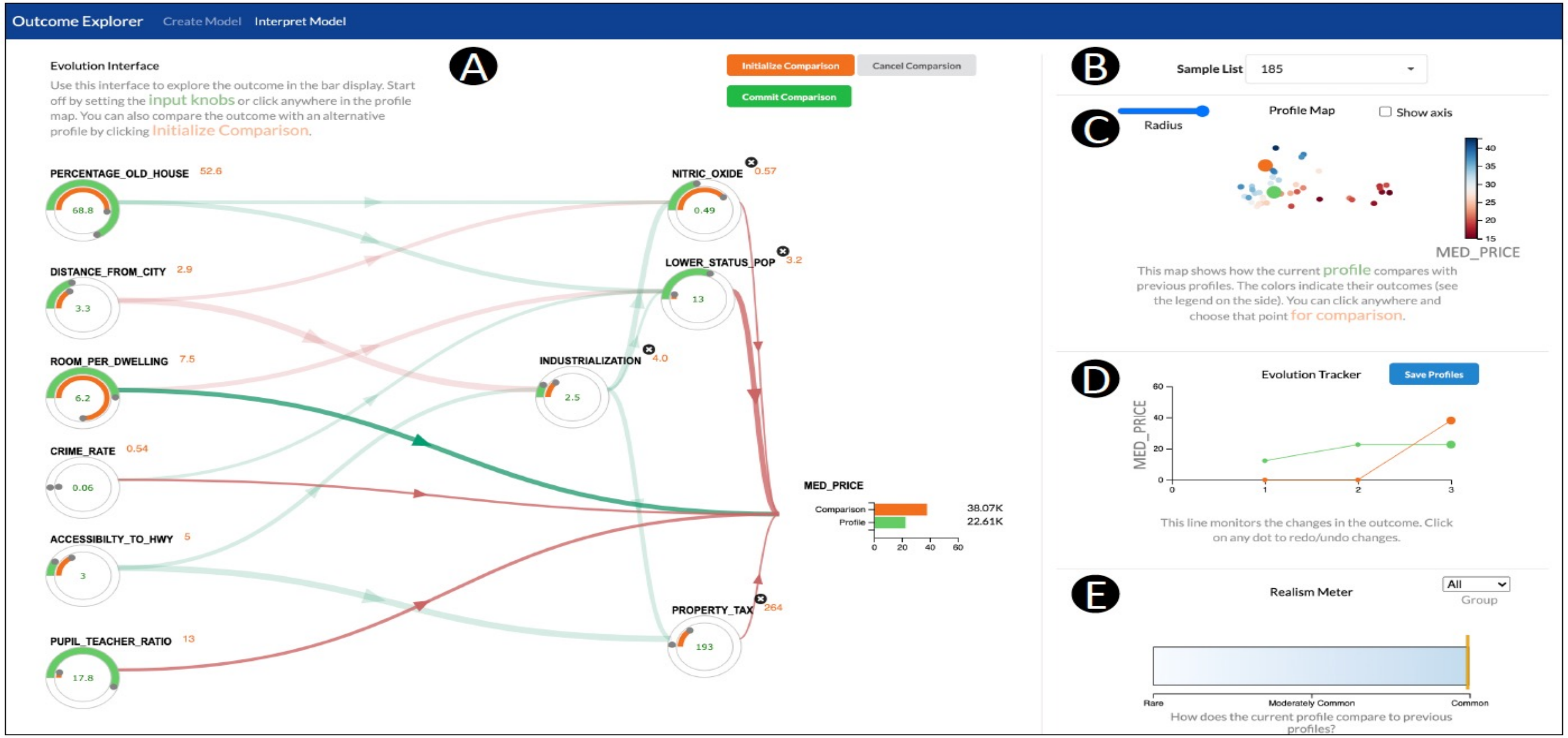
lval	op	rval	estima...	Std. Err.	p value	z-value
NITRI...	~	PERCE...	0.2549...	0.0358...	1.1048... 12	7.1167...
NITRI...	~	DISTA...	-0.	0.0421...	2.0672... 13	-7.344...
NITRI...	~	INDUS...	0.3802...	0.0336...	0	11.290...
LOWE...	~	PERCE...	0.3461...	0.0328...	0	10.524...
LOWE...	~	ROOM...	-0.437...	0.0277...	0	-15.76...

[1]

A) Control Panel. B) Causal structure obtained from the search algorithms. Users can interactively add, remove, and direct edges in the causal structure. C) Model fit measures obtained from Structural Equation Modelling (SEM). D) Parameter estimation for each relation (beta coefficients). E) A line chart showing the prediction accuracy of the Causal Model on the test set.

1. IEEE TRANSACTION ON VISUALIZATION AND COMPUTER GRAPHICS, 2021. AUTHOR'S VERSION. HTTPS://DOI.ORG/10.1109/TVCG.2021.3102051

Method I - Visual Interface



Evolution Interface

Use this interface to explore the outcome in the bar display. Start off by setting the input knobs or click anywhere in the profile map. You can also compare the outcome with an alternative profile by clicking Initialize Comparison.

This map shows how the current profile compares with previous profiles. The colors indicate their outcomes (see the legend on the side). You can click anywhere and choose that point for comparison.

This line monitors the changes in the outcome. Click on any dot to redo/undo changes.

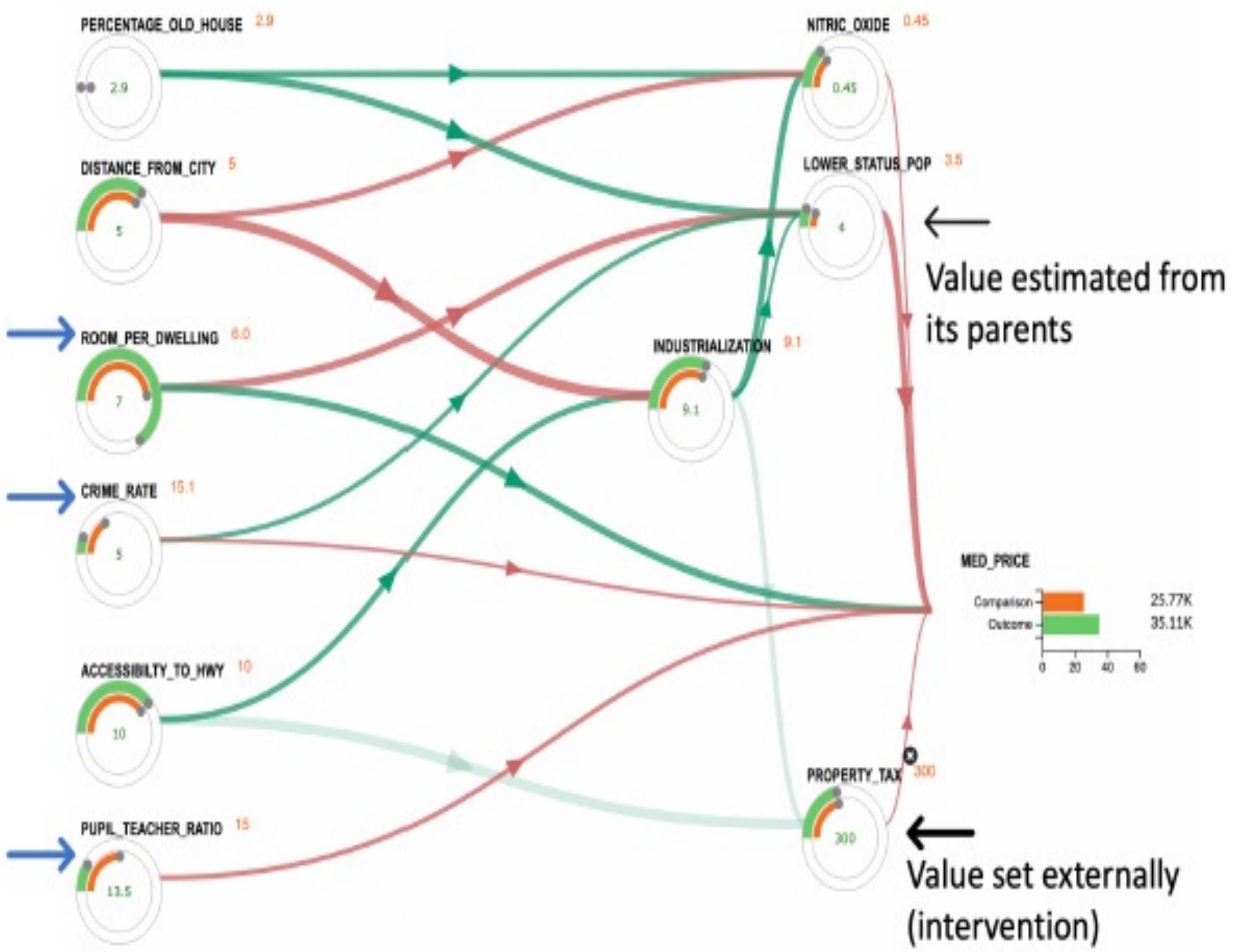
How does the current profile compare to previous profiles?

[1]

A) Interactive causal DAG showing causal relations between variables. Each node includes two circular knobs (green and orange) to facilitate profile comparisons. The edge thickness and colour depict the effect size and type of each edge. B) Sample selection panel. C) A biplot showing the position of green and orange profiles compared to nearest neighbours. D) A line chart to track the model outcome and to go back and forth between feature configuration. E) Realism meter allowing users to determine how common a profile is compared to other samples in the dataset.

[1] IEEE TRANSACTION ON VISUALIZATION AND COMPUTER GRAPHICS, 2021. AUTHOR'S VERSION. HTTPS://DOI.ORG/10.1109/TVCG.2021.3102051

Method I – Exploring Outcomes and What-if Analysis



Value estimated from its parents

Value set externally (intervention)

[1]

- A user can keep one profile (green) fixed, and change the other profile (orange) to ask what-if questions.
- The blue arrows indicate the changes in the orange profile.
- Value is set externally by the user to conduct the analysis.
- As a result, changing the parents will not affect external set value.
- The other endogenous variables are estimated from their parents.

1. IEEE TRANSACTION ON VISUALIZATION AND COMPUTER GRAPHICS, 2021. AUTHOR'S VERSION. HTTPS://DOI.ORG/10.1109/TVCG.2021.3102051

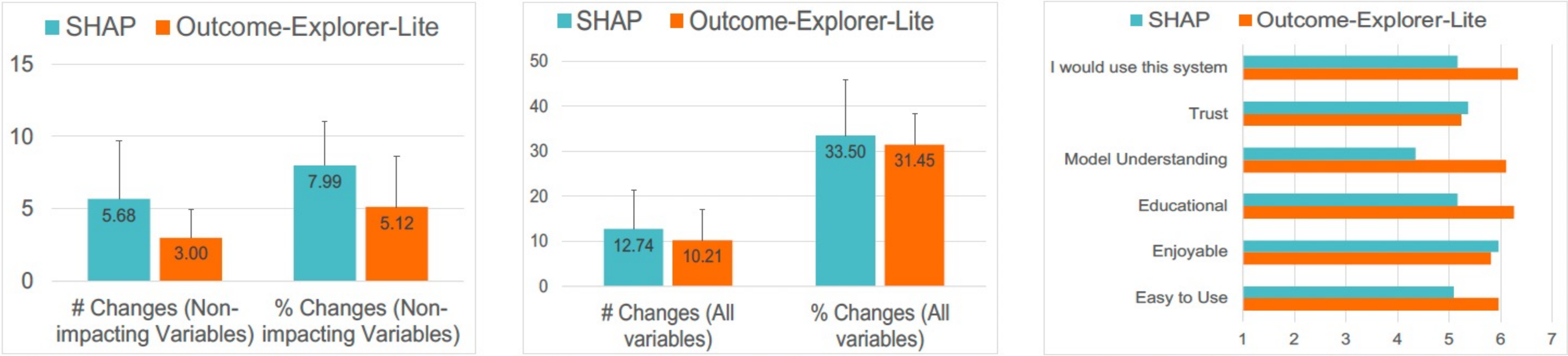
Method I – Evaluation and Results (Expert users):

- Comprehensive and Generalizable - found the accuracy statistics to be most helpful as that feature is not available on other comparable causal analysis tools.

- Engaging, Thought Provoking, and Fun - found the visual design of the Interpretation Module to be aesthetically pleasing and fun to interact with. They mentioned that the interface has a “certain gaming flavour” to it.

- Prior Knowledge and Position in the ML Pipeline - they suggested that Outcome-Explorer could be used once an expert user has pre-processed and explored the dataset. It would provide users the necessary background knowledge for creating and explaining the causal model.

Method I – Evaluation and Results (Non-Expert Users):



(a) Average number of changes and magnitude of changes on non-impacting variables

(b) Average number of changes and magnitude of changes on all variables

(c) Subjective Measures (1: Strongly Disagree, 7:Strongly Agree).

[1]

Study Results. The average number of changes and the average magnitude of changes (%) made to (a) non-impacting variables, and (b) all variables to reach the target outcomes. (c) Average self-reported subjective measures. Error bars show +1 SD.

- C1. SHAP - SHapley Additive exPlanations the state of the art in Machine Learning explainability.
- Outcome-Explorer-Lite: This prototype included only the interactive causal DAG of the Interpretation Module with other components hidden.

[1] IEEE TRANSACTION ON VISUALIZATION AND COMPUTER GRAPHICS, 2021. AUTHOR'S VERSION. HTTPS://DOI.ORG/10.1109/TVCG.2021.3102051

Results from Non-Expert Users

- Outcome-Explorer will increase a user's efficiency in reaching a desired outcome in comparison to the state-of-the-art explanation method.

- Outcome-Explorer will be easy to use.

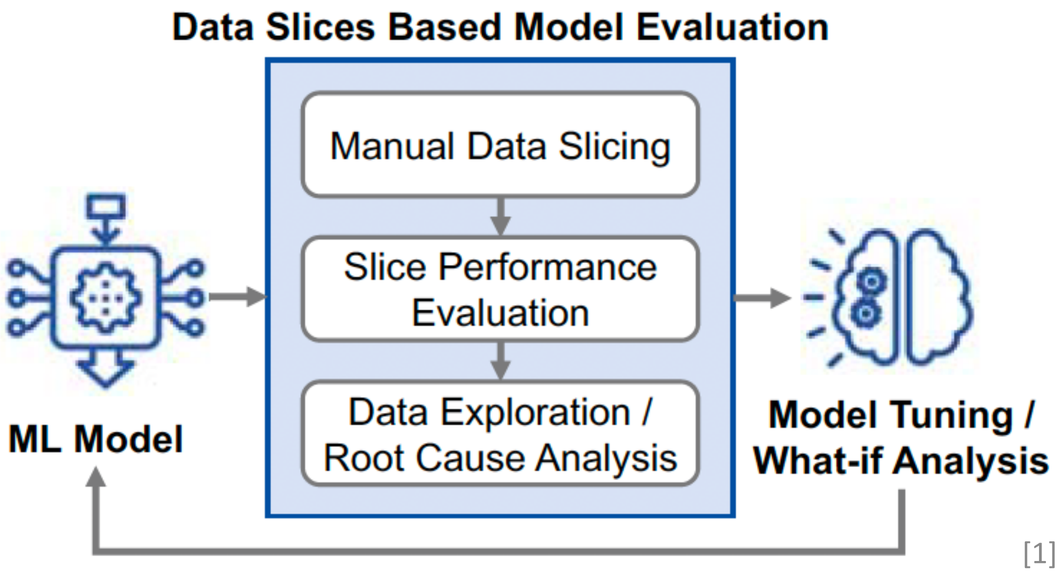
- It will improve user model understanding and that they will learn more about the prediction mechanism using our tool.

- Outcome-Explorer will improve a user's understanding of the embedded predictive causal model in comparison to the state-of-the-art explanation method(SHAP).

Method II - Background

Model improvements based on these reviews:

- Slicing the data is time consuming based on several conditions.
- Users need to explore the related data and model before explaining the scenarios.
- System should enable quick experimentation with the slice-based model.
- Allowing comparison of model performances at slice level.



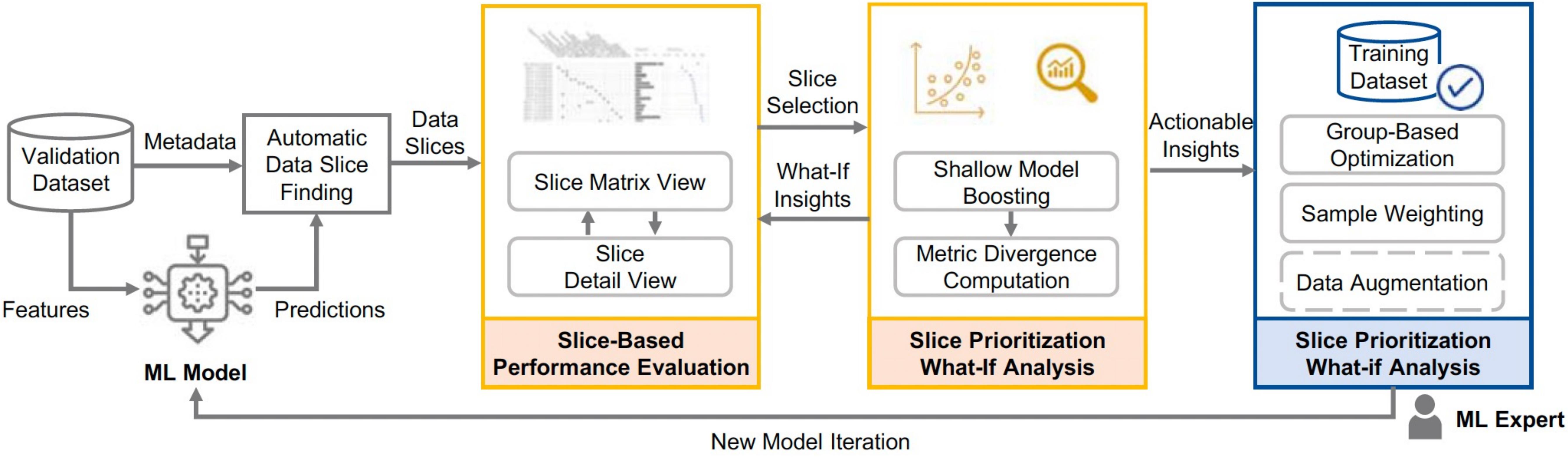
Data Slices Based Model Evaluation
Manual Data Slicing
Slice Performance Evaluation
Data Exploration / Root Cause Analysis
ML Model
Model Tuning / What-if Analysis

[1]

Product MLOps engineer's workflow

1. ZHANG ET AL.: SliceTeller: A Data Slice-Driven Approach for Machine Learning Model Validation

Method II – System Workflow



[1]

Workflow of the model analysis and improvement using SliceTeller . The validation data, together with the model predictions, are used for the automatic slice identification. The produced data slices can be explored using our VA solution (Slice Matrix and Slice Detail View). Users can prioritise groups of data slices and quickly evaluate the effect of this action on the rest of the model slices. Finally, experts can use the insights gained from the system to fine tune the model and continue the analysis with SliceTeller .

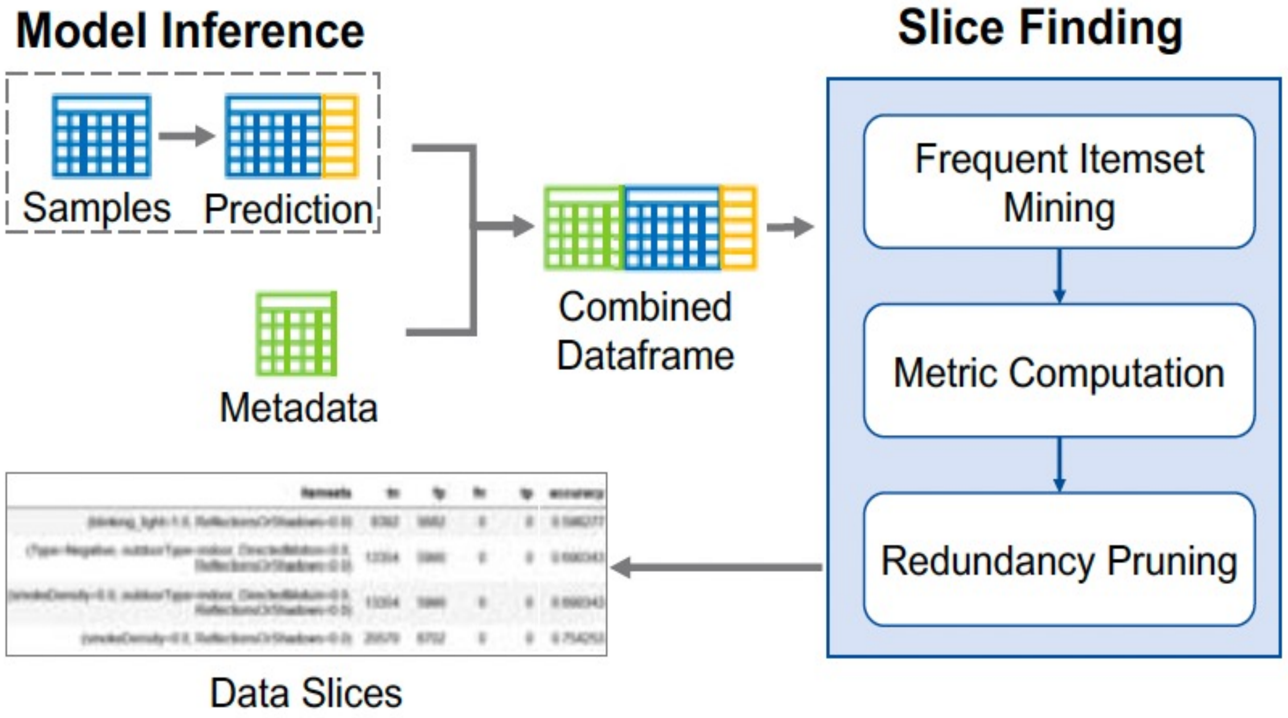
1. ZHANG ET AL.: SliceTeller: A Data Slice-Driven Approach for Machine Learning Model Validation

Method II – Data Slicing (DivExplorer Algorithm)

- The algorithm takes the model predictions and the meta-data (interpretable features of the dataset) as input and executes an exhaustive slice search by frequent pattern mining (approach to remove data which is not meaningful or unique to optimise performance).
- Algorithm is run twice

  - Firstly, to identify the relevant metadata features which are most correlated with poor performance.

  - Secondly, to perform a fine-grained search (parameters can be fine tuned by users).

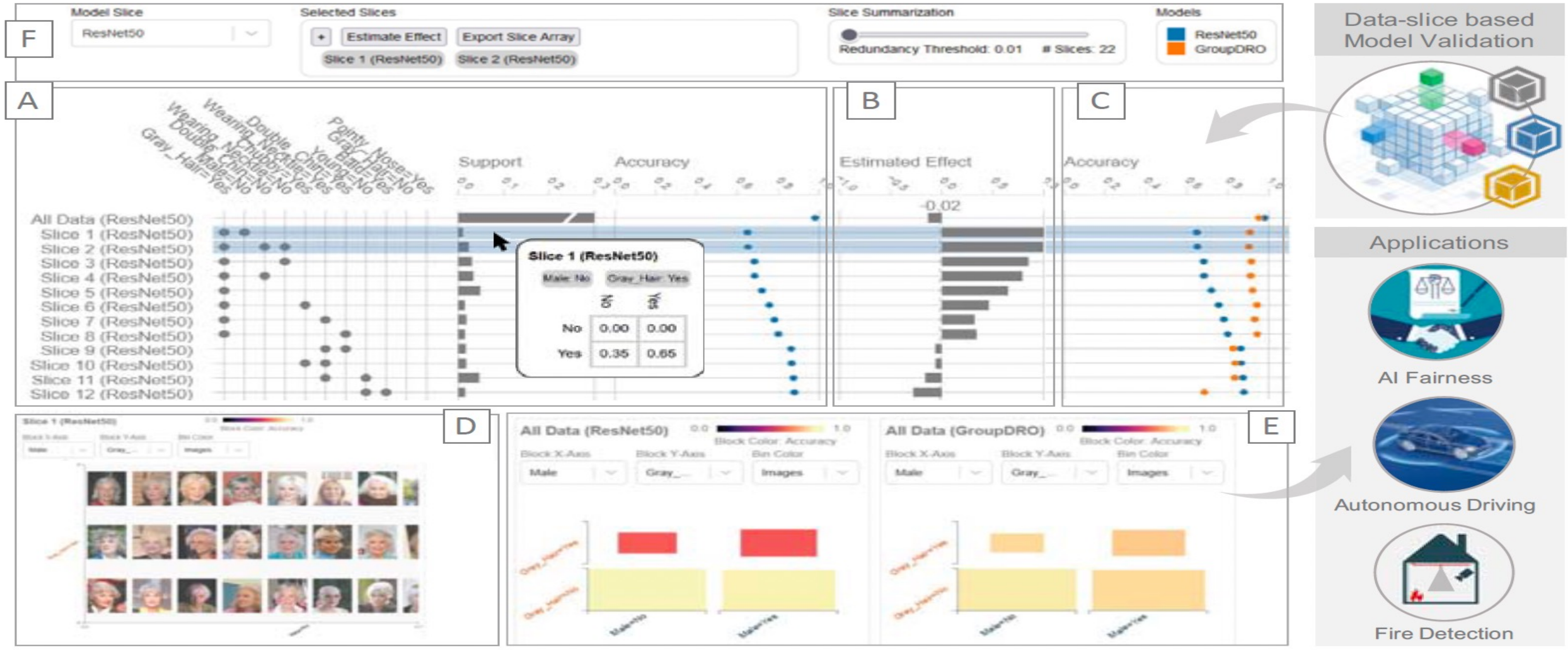


itemsets	tn	fp	fn	tp	accuracy
(blinking_light=1.0, ReflectionsOrShadows=0.0)	8362	5552	0	0	0.596277
(Type=Negative, outdoorType=indoor, DirectedMotion=0.0, ReflectionsOrShadows=0.0)	13304	5866	0	0	0.698043
(smokeDensity=0.0, outdoorType=indoor, DirectedMotion=0.0, ReflectionsOrShadows=0.0)	13304	5866	0	0	0.698043
(smokeDensity=0.0, ReflectionsOrShadows=0.0)	20570	6702	0	0	0.754252

[1]

1. ZHANG ET AL.: SliceTeller: A Data Slice-Driven Approach for Machine Learning Model Validation

Method II – Design



[1]

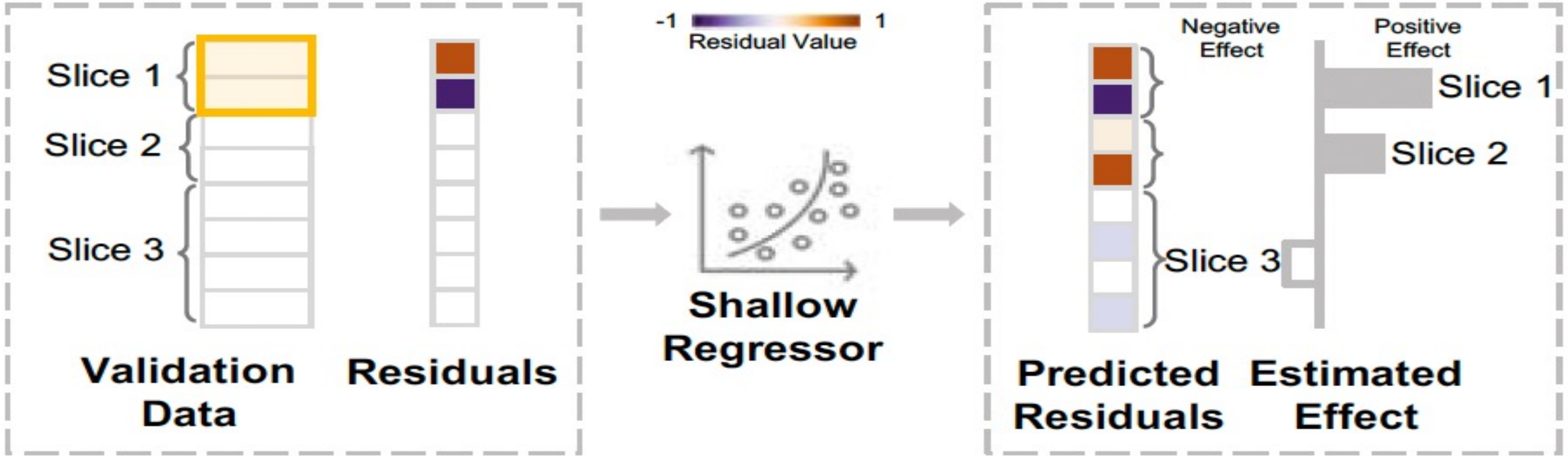
(A) Slice Matrix: The data slices (represented as rows), slice descriptions (encoded as columns), and slice metrics (Support and Accuracy). Slices are sorted by model accuracy. (A - Tooltip) Confusion matrix for Slice 1. (B) Estimated effects of optimising the model for two data slices (Slices 1 and 2, highlighted in blue). (C) Accuracy comparison between the two models, ResNet50 and GroupDRO. (D) Slice Detail View containing image samples from a data slice. (E) Slice Detail View containing the comparison of two data slices using the MatrixScape visualisation. (F) System menu, containing options for model selection, effect estimation of focusing on a slice during model training, and data slice summarization.

VISUAL COMPUTING GROUP HEIDELBERG UNIVERSITY

1. ZHANG ET AL.: SliceTeller: A Data Slice-Driven Approach for Machine Learning Model Validation

Method II – SliceBoosting Algorithm

The main idea is that instead of training the full model to evaluate slice trade-offs, we can train a shallow model to approximate the residuals (errors) of the slices

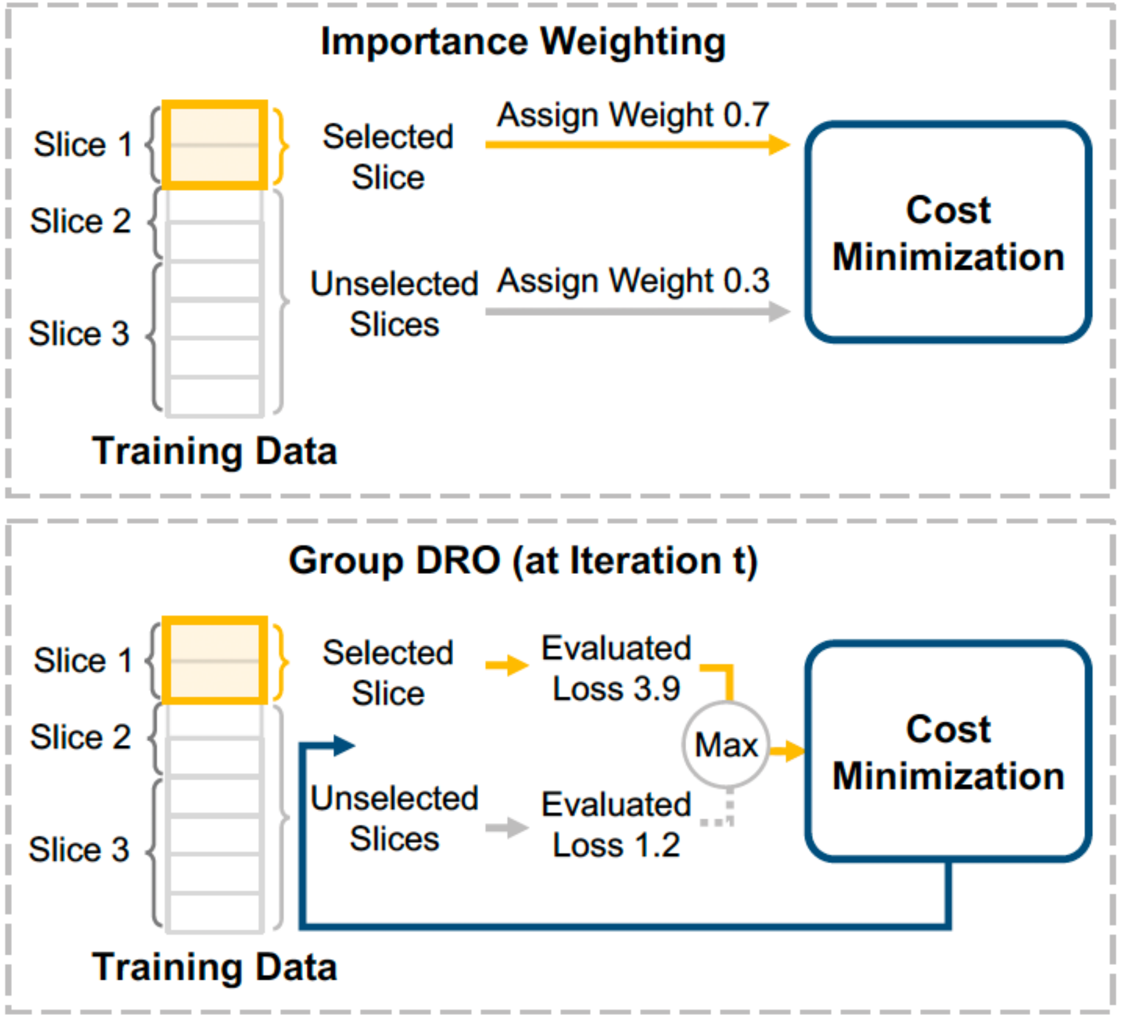


[1]

- Given the selected slice 1, shallow regressor is trained to estimate that, under the ideal scenario where the optimization model correctly fits to slice 1:
  - How will this effect the performance of slice 1 and 2.
- Prediction target of the regressor is designed as the residuals of the original model predictions.
- To focus on the effect estimation of slice 1, the residual values are set only for slice 1, while keeping the residuals of all other slices as 0.
- The predicted residuals from the regression are in the range of [−1,1].
- Aggregation of the sample-level residual predictions to obtain slice-level estimation results.

1. ZHANG ET AL.: SliceTeller: A Data Slice-Driven Approach for Machine Learning Model Validation

Method II – Model Optimisation

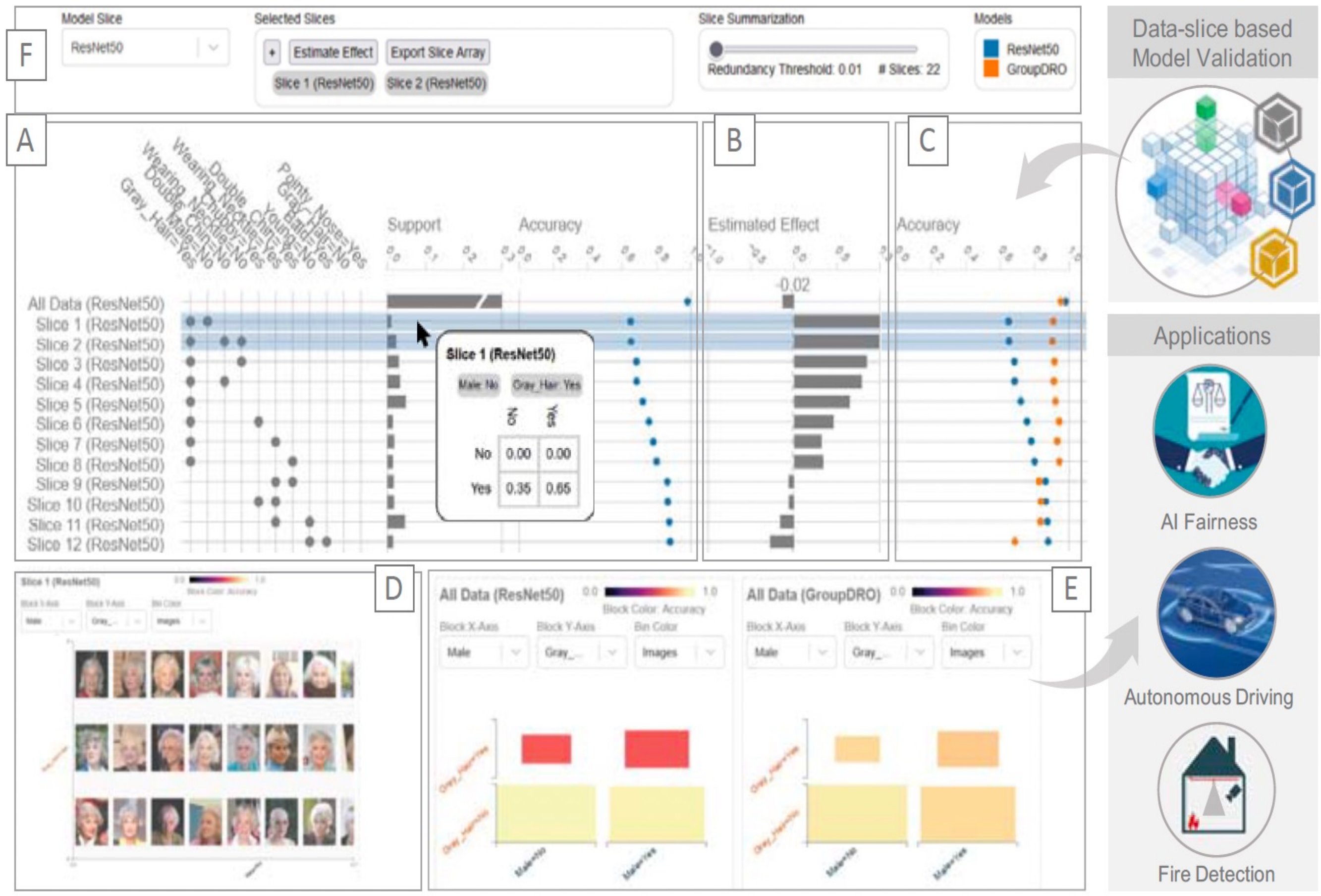


[1]

- Importance weighting method changes the loss function by assigning heavier weights to the training samples in the worst-performing slices.
- Group Distributionally robust optimization prioritises the worst-performing slices during the training process.
- During re-training, the model prioritise slices in the training data according to user's decision.

1. ZHANG ET AL.: SliceTeller: A Data Slice-Driven Approach for Machine Learning Model Validation

Method II – Evaluation and Results

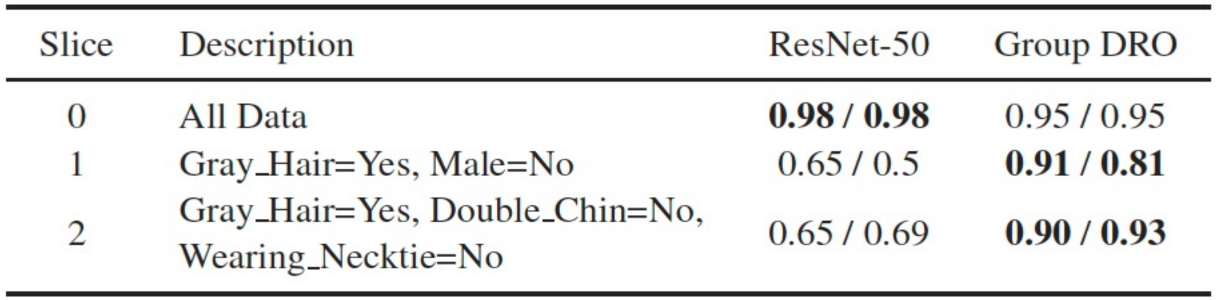


[2]

Case I – Bias Detection for AI Fairness in Image Classification Models

- The CelebA dataset contains 202,599 face images of 10,177 celebrities, along with 40 binary attribute annotations including gender, skin color, smiling, etc. for each image.
- Splitting of Data: training(70%), validation(20%), testing(10%).
- 1 – Data is trained on slice 1 and 2 to check the accuracy on ResNet.
- 2 – Data is retrained on GroupDRO with Slice 1 and Slice 2 together to get a better accuracy.

Table 2. Accuracy of Image Classification Models (Val / Test)

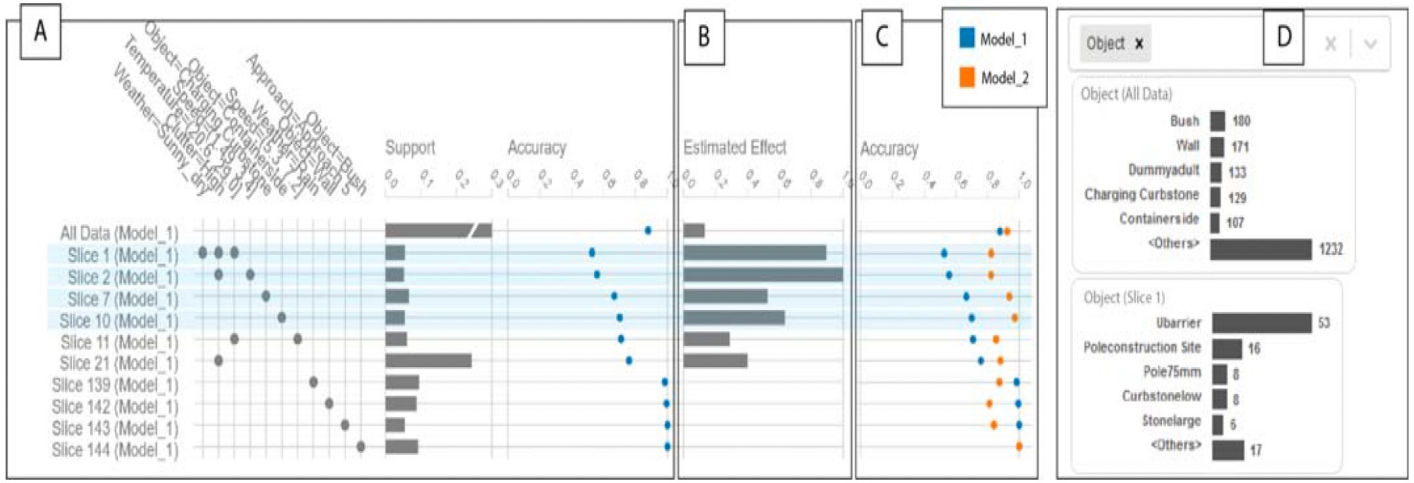


Slice	Description	ResNet-50	Group DRO
0	All Data	**0.98 / 0.98**	0.95 / 0.95
1	Gray_Hair=Yes, Male=No	0.65 / 0.5	**0.91 / 0.81**
2	Gray_Hair=Yes, Double_Chin=No, Wearing_Necktie=No	0.65 / 0.69	**0.90 / 0.93**

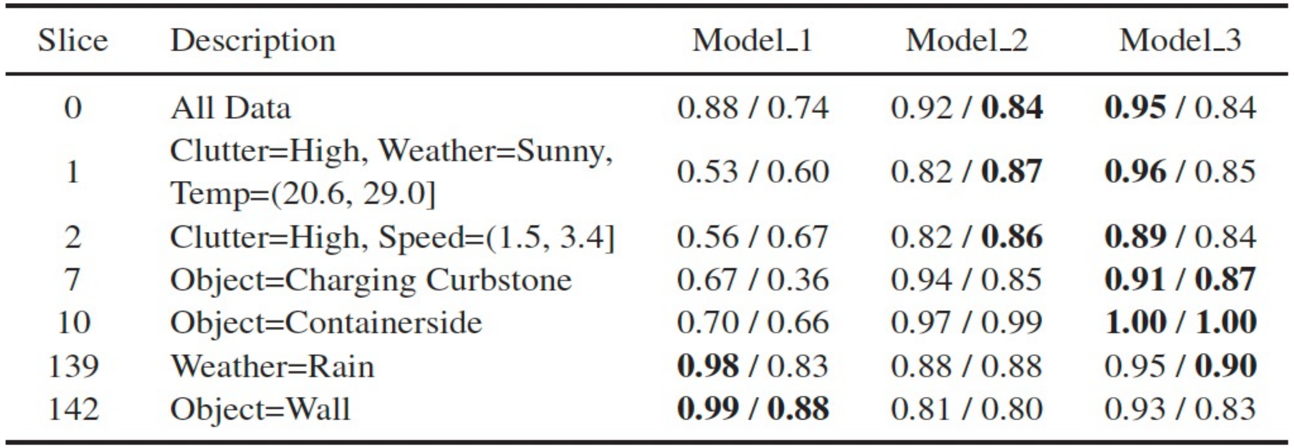
[1]

1. ZHANG ET AL.: SliceTeller: A Data Slice-Driven Approach for Machine Learning Model Validation
2. ZHANG ET AL.: SliceTeller: A Data Slice-Driven Approach for Machine Learning Model Validation

Method II – Evaluation and Results



[1]



Slice	Description	Model_1	Model_2	Model_3
0	All Data	0.88 / 0.74	0.92 / **0.84**	**0.95** / 0.84
1	Clutter=High, Weather=Sunny, Temp=(20.6, 29.0]	0.53 / 0.60	0.82 / **0.87**	**0.96** / 0.85
2	Clutter=High, Speed=(1.5, 3.4]	0.56 / 0.67	0.82 / **0.86**	**0.89** / 0.84
7	Object=Charging Curbstone	0.67 / 0.36	0.94 / 0.85	**0.91** / **0.87**
10	Object=Containerside	0.70 / 0.66	0.97 / 0.99	**1.00** / **1.00**
139	Weather=Rain	**0.98** / 0.83	0.88 / 0.88	0.95 / **0.90**
142	Object=Wall	**0.99** / **0.88**	0.81 / 0.80	0.93 / 0.83

[2]

Case II – Ultrasonic Object Height Classification for Autonomous Driving

- The dataset has 157,743 records that consists of 71 numerical features, and the expert's goal was to predict object height as a binary label: 'high' or 'low'.
- Training the model using XGBoost Algo by Splitting of Data: training(60%), validation(30%), testing(10%).
- All Data is trained on checking the accuracy on Model_1
- After evaluation the slice matrix and effects, they train slice 1,2,7, 10 to get a better accuracy.
- Slice trade-off was noticed with slices 139,142 and 143 with a significant drop.
- Model_3 has better results on all the slices than model_1 where the model performs the worst.

1. ZHANG ET AL.: SliceTeller: A Data Slice-Driven Approach for Machine Learning Model Validation
2. ZHANG ET AL.: SliceTeller: A Data Slice-Driven Approach for Machine Learning Model Validation

Expert Reviews

- The automatic data slicing and the matrix view reduces the model testing time and effort.
- The summarization slider and the matrix view could facilitate the model exploration and help reduce the amount of time they needed to spend looking at data slices.
- SliceTeller is interpretable and with a detailed visual approach, it will fasten the tasks for ML engineers to analyze the data with the flaws and take immediate actions on it.

[1]

1. https://www.shutterstock.com/search/machine-learning-expert

Comparison

Outcome Explorer

- Outcome Explorer focuses on expert as well as non-expert users.
- It is a decision making tool which provides the detailed explanation of the algorithm in the form of causal guided interface.
- It helps in the application areas and allows non-expert users to make decisions on their own by letting them compare their result with several other results in the same field.
- Ex – doc can compare the disease of a patient and predict the outcome based on the similar side effects of other patients.

SliceTeller

- Slice Teller focuses only on expert users.
- It allows to handle a large data set by breaking them into smaller slices for evaluation
- It allows the expert users to conclude at a faster time due to details information of the data slices.
- Smaller slices of data helps in identifying and correcting errors.
- Ex – model can assist in image-based fire detection where there is a potential of identifying fires at an early stage

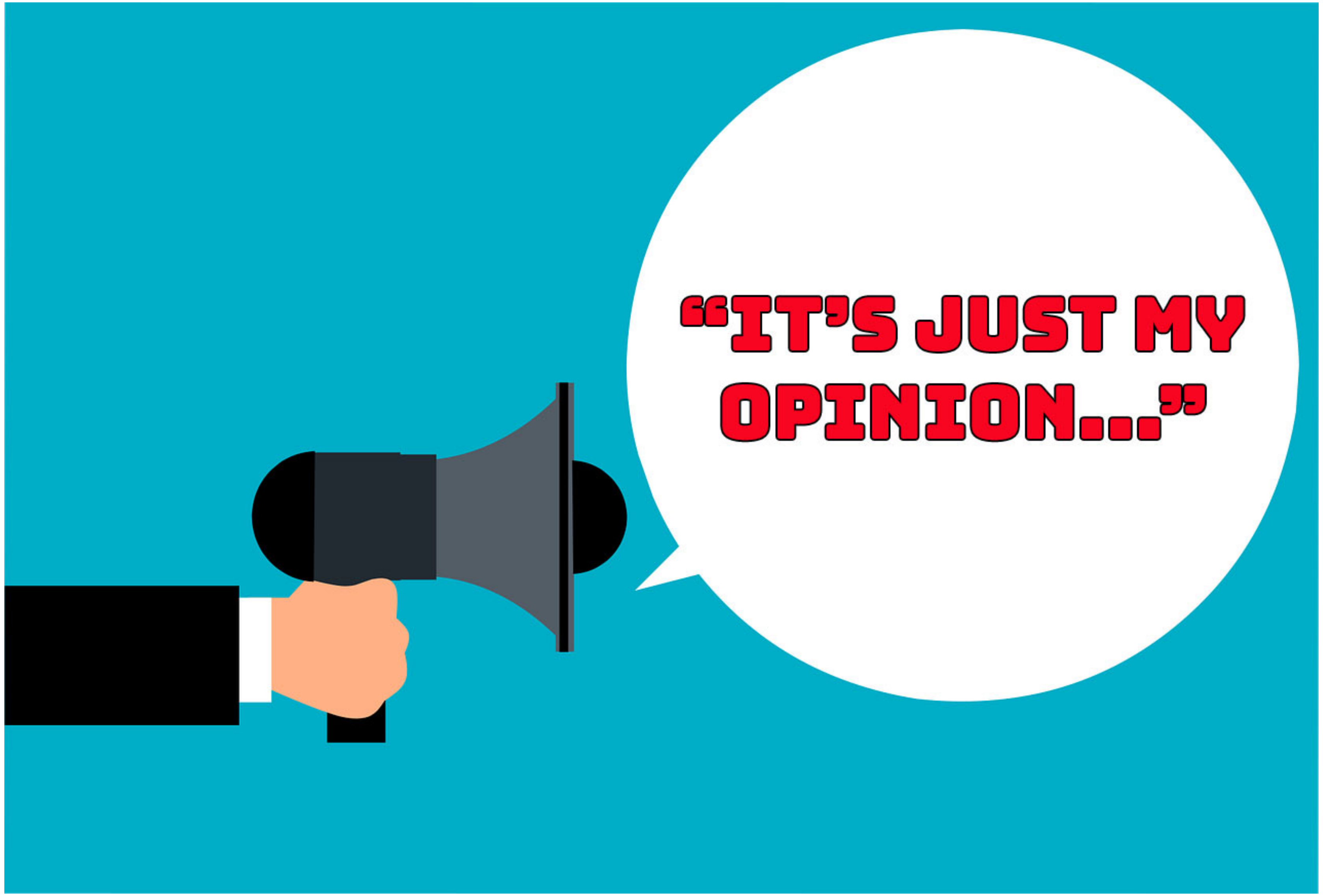
Limitations – Outcome Explorer

- There are several disadvantages of using outcome explorer, a causality-guided interactive visual interface for interpretable algorithmic decision making:
- Computational Intensity: Requires large sample size to produce robust results.
- Causality Assumptions: Makes assumptions about causality that may not hold in all cases, causing inaccurate or misleading results.
- Limited interpretability: Outcome explorer may be difficult to interpret for non-technical users, which can limit its usefulness in certain settings.

Limitations - SliceTeller

- Limited Generalization: A model that is only validated on a specific subset of data may not perform well on new, unseen data.
  - This is because the model may have overfitted to the specific characteristics of the data slice it was trained on.
- Lack of Representation: Validation on a small data slice may not give a good representation of the entire dataset.
  - This could lead to the model being under- or over-performing on unseen data.
- Lack of Data: Slicing the data can lead to having too small of a dataset to work with.
  - It can lead to poor model performance, as well as lack of confidence in the results.

My Opinion



[1]

1. https://myburbank.com/opinion-it-is-time-to-add-a-new-subject-to-the-talks//

Conclusion

- Data slicing can be efficient for quickly analysing specific subsets of data, such as specific groups or time periods.
  - However, one drawback of data slicing is that it can lead to a loss of context or a biased view of the data if not used carefully.
- Outcome can be efficient for identifying potential causal relationships and understanding the factors that influence a particular outcome.
  - However, one drawback of outcome explorer is that it can be computationally intensive and requires a large sample size to produce robust results.
- In general, both data slicing and outcome explorer have their own advantages and drawbacks.
  - The best method to use will depend on the specific goals and characteristics of the dataset being analysed.

Thank you for your attention!
Any Questions?