How Would Visualization Help Enhancing Trust into Machine Learning Models?

Description

Thank you for agreeing to take part in ISOVIS (http://cs.lnu.se/isovis/) group's anonymous survey!

A good understanding of why a Machine Learning (ML) model produces a result is crucial, especially, if one wants to make a decision based on such results. Therefore, it is important for an analyst to trust the model that she/he wants to deploy. Visualization of various aspects of ML models has been used to address this problem at various degrees of success.

The aim of this survey is to understand what step(s) of the entire computational process or specific properties of the ML models and of the data could be visualized with the goal of enhancing trust into the chosen model.

We will not gather any information that could be used to track you personally. Nevertheless, we will do our best to keep your information confidential if you choose to disclose any of your own free will.

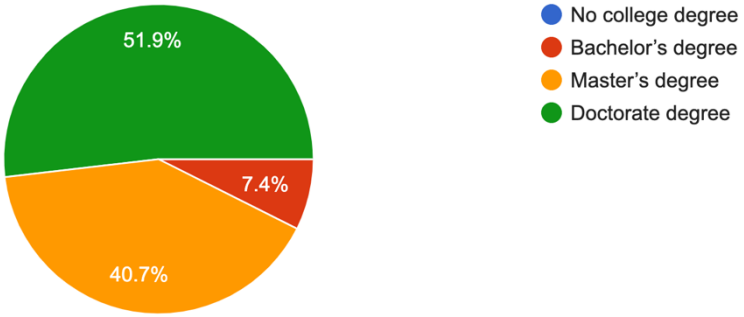
Survey time: approximately 10 minutes

General Statistics - Statistical Information

Note that in some of the following questions, we intentionally excluded 'Reinforcement Learning' from the types of ML.

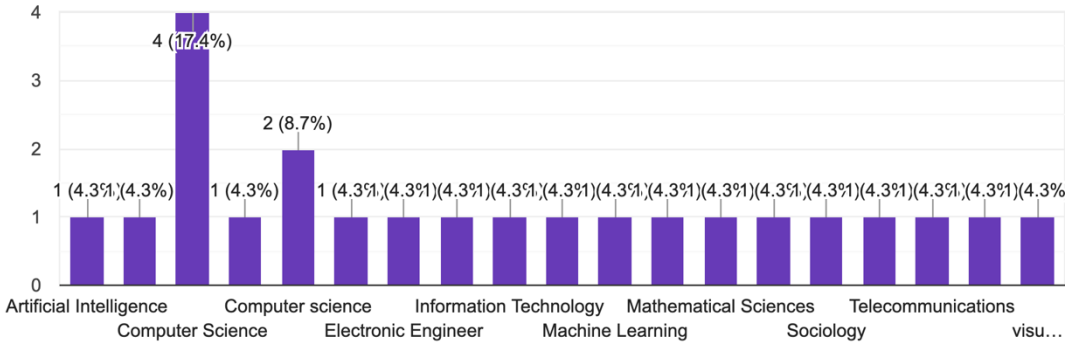
Your education

27 responses



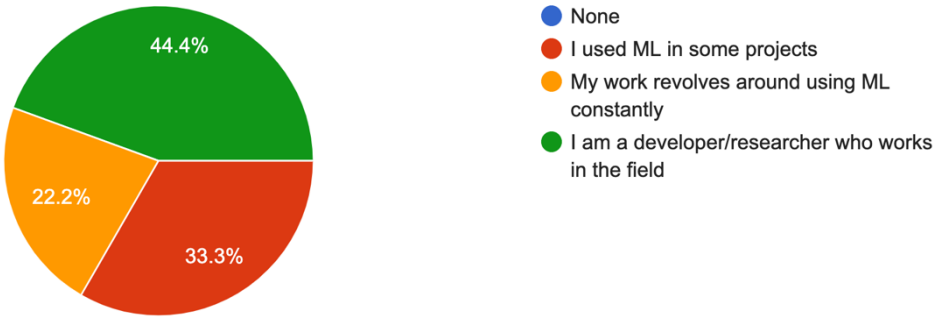
If you previously selected having a finished degree, then in which subject?

23 responses



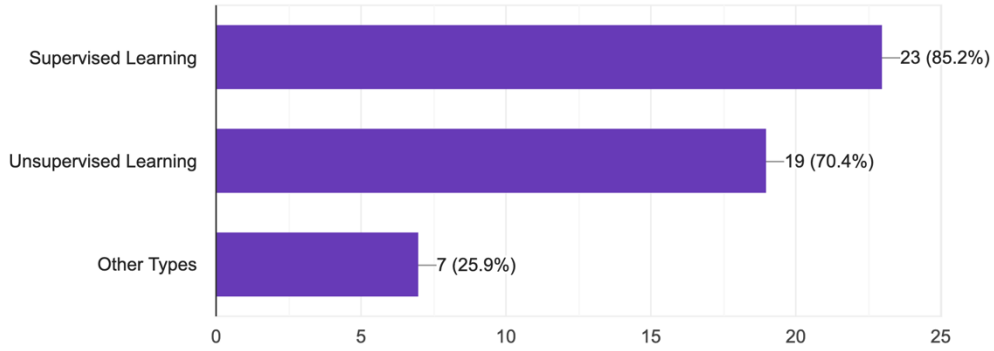
Your experience using ML in your work/life?

27 responses



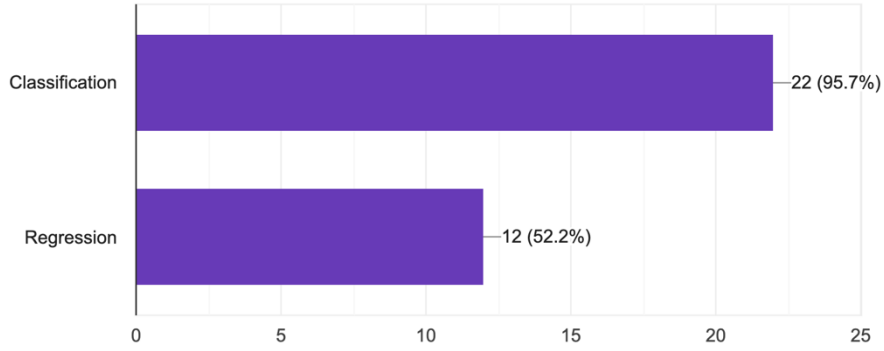
What types of ML algorithms/models have you used the most?

27 responses



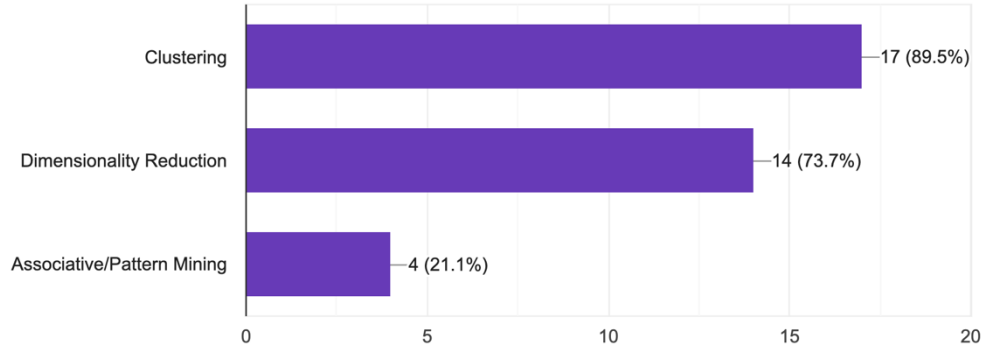
If you previously answered Supervised Learning, then which specific type(s)?

23 responses



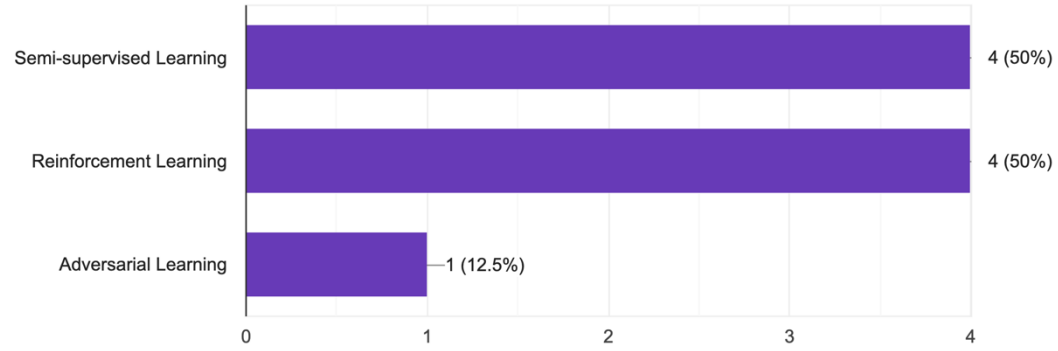
If you previously answered Unsupervised Learning, then which specific type(s)?

19 responses



If previously answered Other Types, then which specific type(s)?

8 responses



A Hypothetical Application Scenario

A typical ML process requires the following steps: (1) data preparation, (2) training step, (3) evaluation step, (4) production deployment (i.e., predicting according to the testing data) and iterating/looping through them.

Let's set up a hypothetical application scenario, where you have to train an ML algorithm such as a neural network or a random forest, without knowing which of them performs best. A well-known data set (https://www.kaggle.com/uciml/pima-indians-diabetes-database) is used for this hypothetical scenario: it is about female patients from Pima Indian heritage, who may or may not have diabetes. The data set consists of several medical predictor variables, such as glucose concentration in blood, insulin level, and blood pressure. In addition, there is one target variable, which shows whether a patient is positive or negative to diabetes. This scenario is used to exemplify the questions in this questionnaire (see the notes in parentheses after each question).

Now, please try to answer—as honestly as possible—the following questions. Note that in case you do not know an answer, you can skip the question as most of them are optional. The Likert scale is from 1 (strong disagreement) to 5 (strong agreement) with 3 being "neutral."

Brief Glossary (Optional)

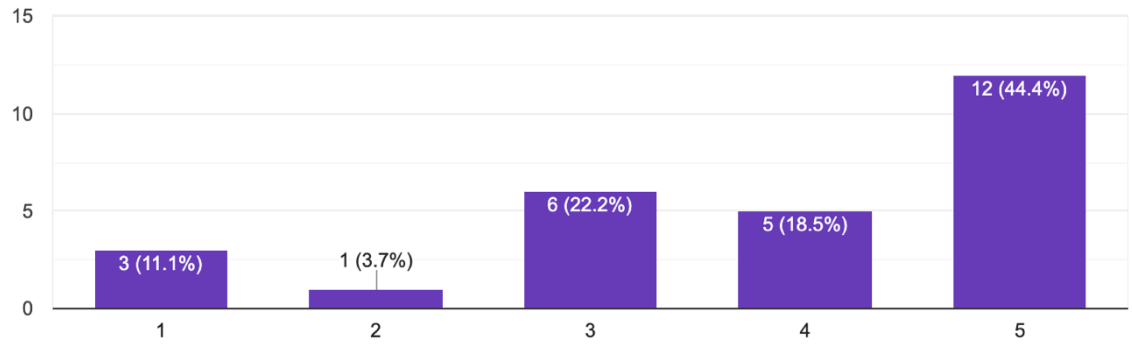
First of all, we present a brief glossary for our survey, which you might want to read if you already do not know about the following terms:

- Steps of the process: the different process steps of a typical ML pipeline (see the paragraph below).
- Algorithms: ML techniques, such as K-Nearest Neighbors or Random Forests, which can have various and diverse parameters.
- Models: the results of the ML algorithms, i.e., concrete classifiers/models with specific sets of parameters (or hyperparameters).
- Random forests: operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) of the individual trees.
- Neural networks: a set of algorithms, modeled loosely after the human brain, that are designed to recognize patterns.
- Instances: (usually) the rows of a table representing the samples of a data set.
- Features: (usually) the columns of a table representing the dimensions of a data set.
- Training data: with which you train your model in order to learn how to predict.
- Testing data: with which you check how well your model performs when receiving new data.

Multiple Choice Question:

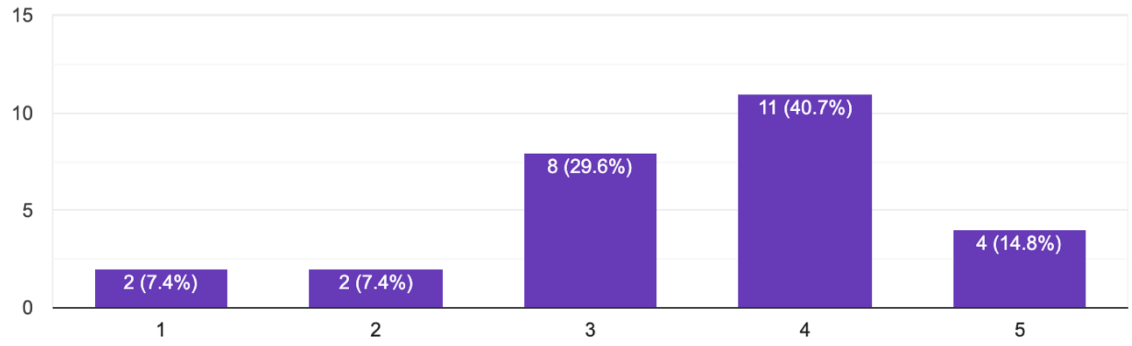
Q1) Details about the source of the data should be visualized (for instance, that researchers and doctors collected the data).

27 responses



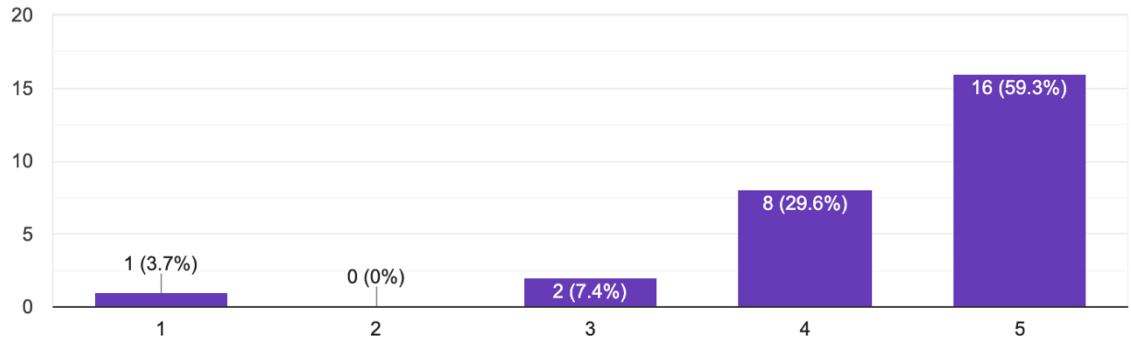
Q2) Details about the data collection process should be visualized (for example, that special equipment measured glucose and insulin values).

27 responses



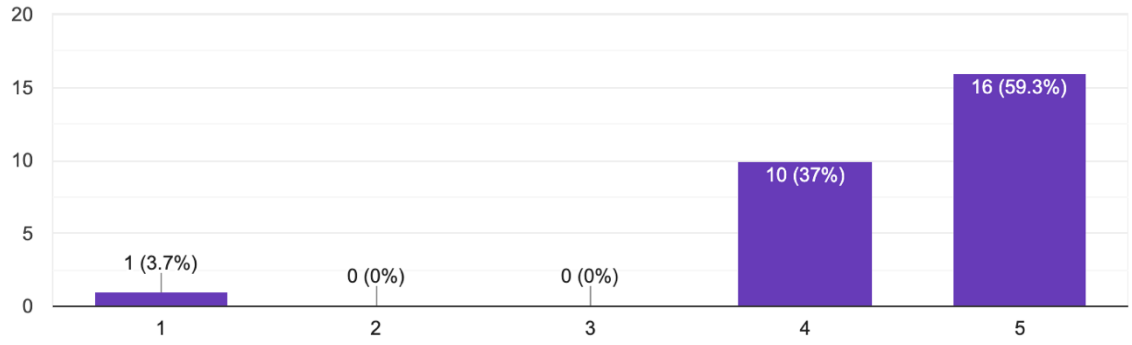
Q3) Any data quality issues should be visualized (e.g., that some insulin values are 0 and for what reason).

27 responses



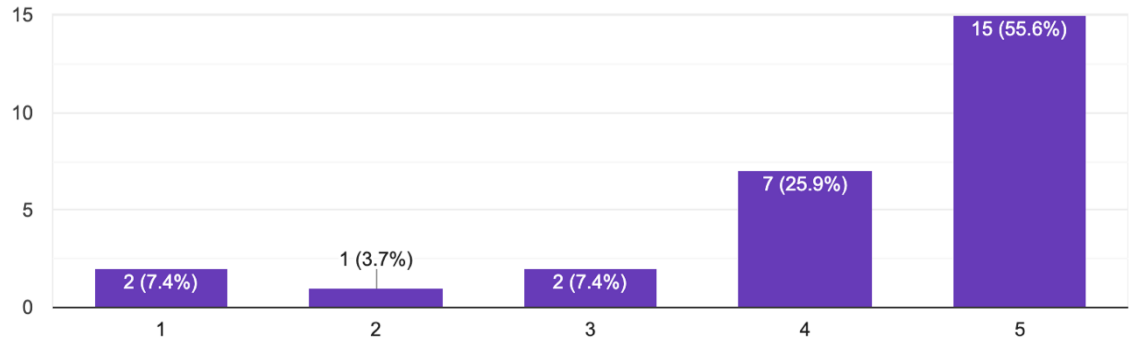
Q4) Performances of the machine learning algorithms should be compared using visualization (as in our scenario where there are several algorithms to choose from).

27 responses



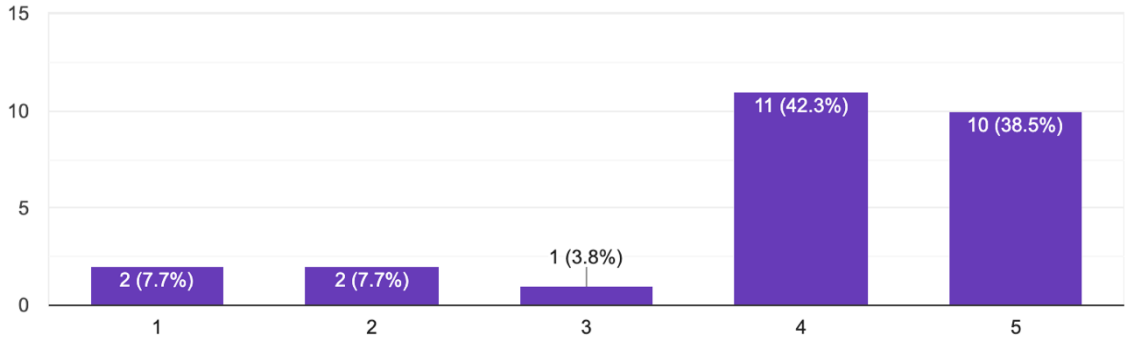
Q5) Tuning parameters (or hyperparameters) and finding the best model with the help of visualization is important (e.g., you chose random forests with 80 or 120 number of estimators/decision trees).

27 responses



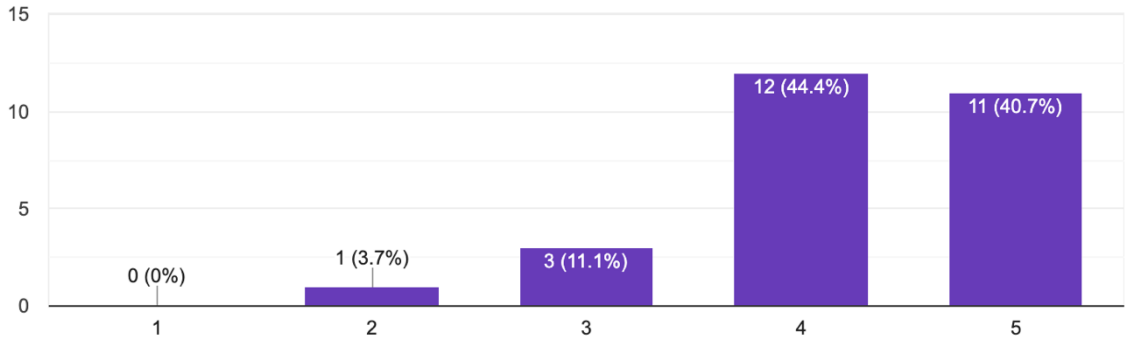
Q6) Gaining control of the data during the training process by using visualization is crucial because you can manipulate data instances and features (by, for example, removing some similar instances that are negative to diabetes in order to balance the data).

26 responses



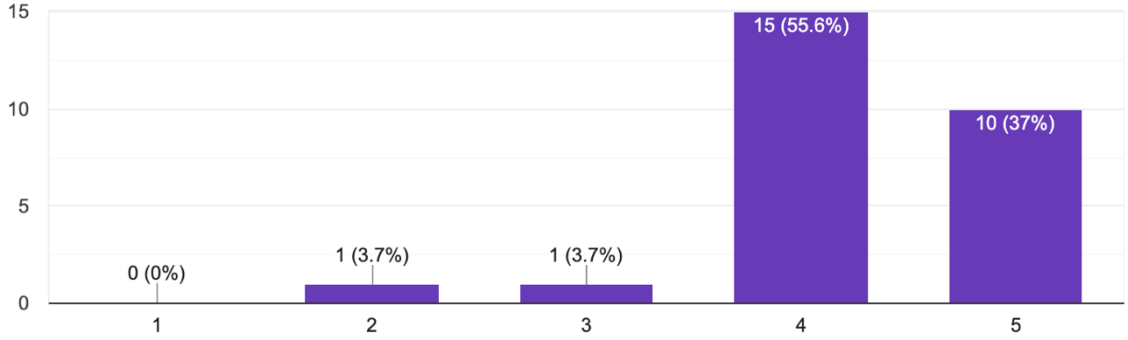
Q7) Trustworthiness is influenced by the ability of the system to visualize automatically feature importance (with respect to the ML task) and feature dependencies (e.g., detecting that glucose and insulin are two highly correlated and necessary features).

27 responses



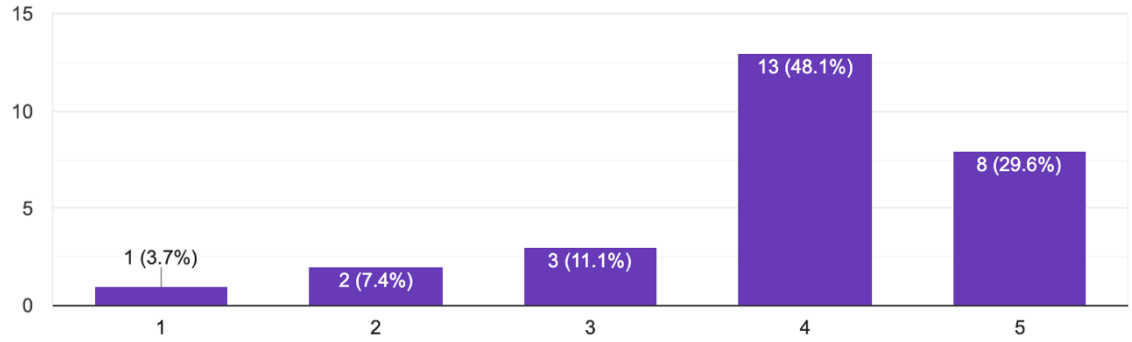
Q8) Visualizing the way the decision of the model is done enhances trustworthiness in the results (for instance, by showing the decisions taken by random forest).

27 responses



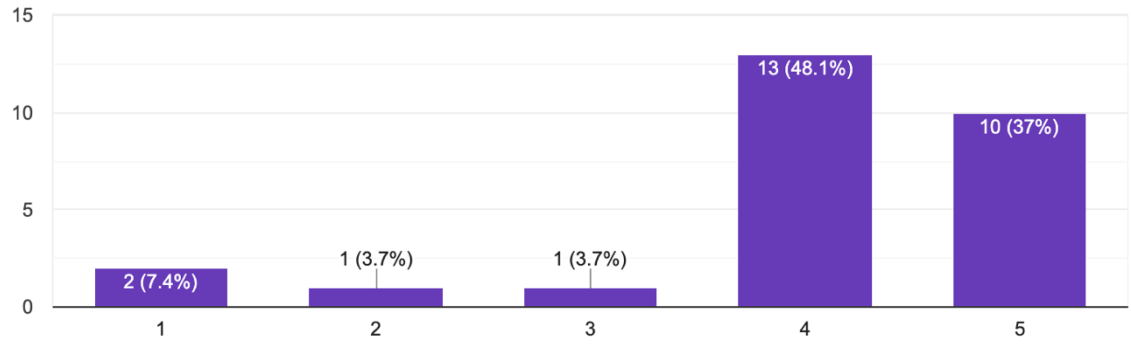
Q9) Steering the learning process with the use of visualization is important (e.g., filtering out some decision trees from random forests).

27 responses



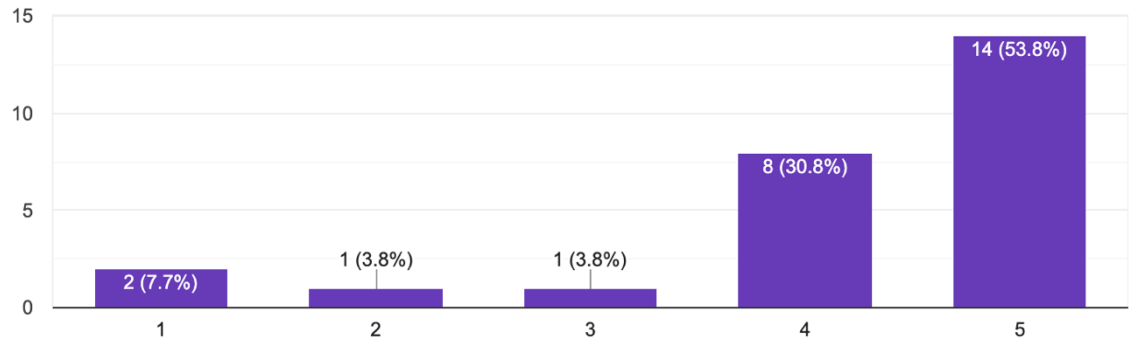
Q10) The different visualization views should automatically guide you to the most "game-changing" decisions of a model (for instance, by highlighting relevant areas).

27 responses



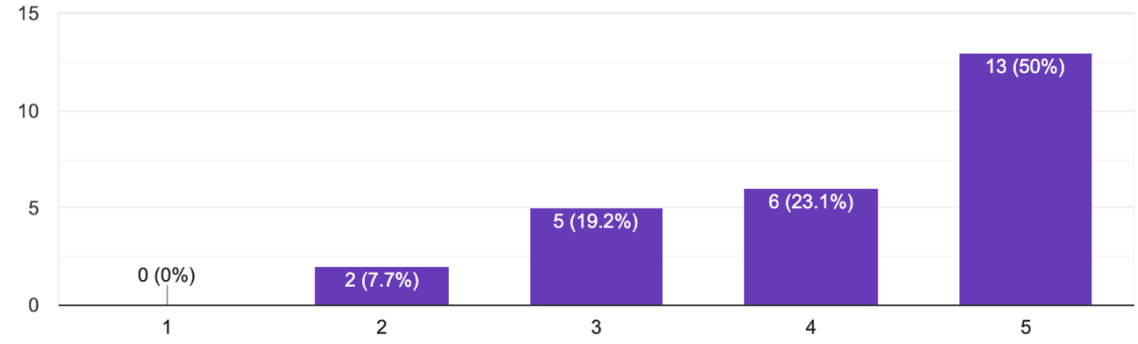
Q11) Visual representations should allow multiple "what-if" scenarios and examining the possible outcomes based on your actions (for example, what if you used specific parameters instead of the ones that you had chosen before).

26 responses



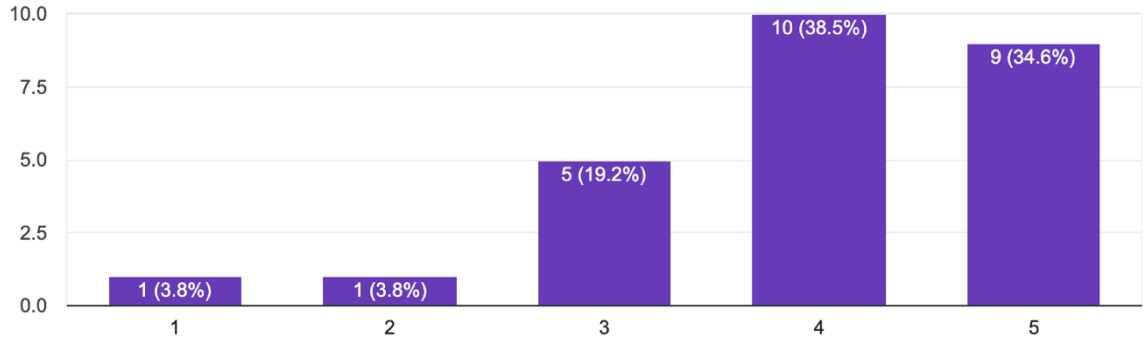
Q12) Using visualization to explore cases that the model did not fairly judge instances is crucial (for instance, if glucose levels are high then the patient automatically is positive to diabetes without taking into account the other features at all).

26 responses



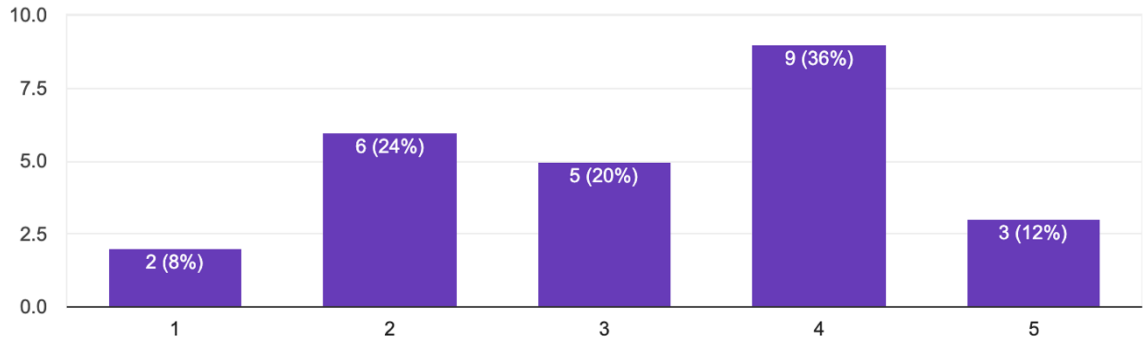
Q13) Visualization should enable collaboration between colleagues by allowing dynamic annotations and letting the experts agree or disagree (for example, by visualizing the agreement rate between experts for insulin and glucose being important features for the classification).

26 responses



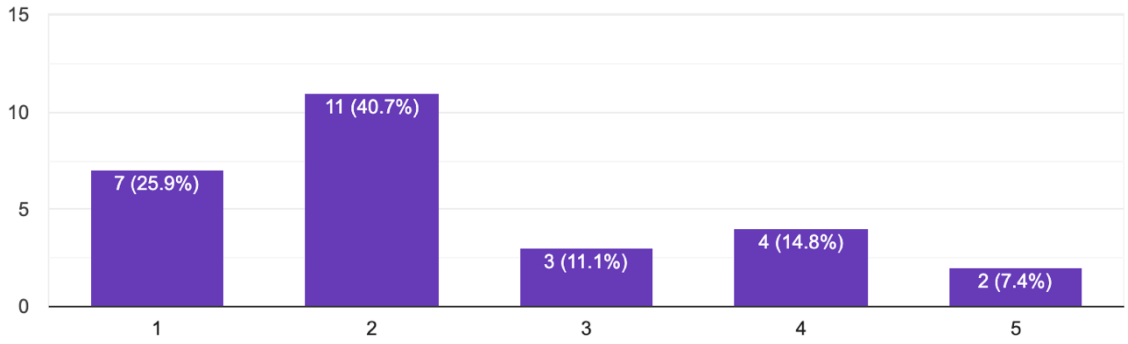
Q14) Choosing an evaluated/user-tested and popular visualization tool (and not others) for a specific ML model (e.g., for random forests) are important factors to you.

25 responses



Q15) When an adapted performance metric shows that your model reaches your expectations, you do not need additional information such as an explanation of the way predictions are made (for instance, if a predictive model reaches 95% of accuracy on a test set in our example).

27 responses



Question:

In your opinion, what step(s) of the process or property/properties of the models and of the data would you like to visualize to increase the trust into the ML model(s) you are using?

Answers (27 responses):

Model decision boundary (as for SVM) for complex models such as NN
Those listed above.
The importance of the features in the classification/prediction; what makes one class different from another (what makes the algorithm put them apart?); uncertainty in the classification/prediction.
Testing model prediction to check fitting of model. Eg., learning curves
Importance of inputs, possibly connection to outputs (in some reasonable format!)
Visualizing the steps for decision making is important. And this can be achieved for example by visualizing the weights (coefficients) of the different features that were used to make a certain decision (class or numerical prediction). In case an outlier is predicted by the model, then the user by looking at visualizations could see which features contributed the most to that decision. This could also be complemented with visualizations with raw data. For instance, I have been working with clinical studies with Parkinson's disease patients wearing sensors in their wrists. For us researchers, it was difficult to see how the data was collected e.g. patients could do a certain daily activity (e.g. cutting grass) but in our model we accounted that as tremor. So having a tool that visualizes the raw data could complement investigation of interesting (outliers) cases in the dataset tracking from decision to actually what happened in the real world.
All design choices in the process that make a substantial difference in the outcome should be made aware to the user.
How hyper-parameters affect the model. How suitable is the dataset for a specific model.
Evaluation step.
Visualisation of pre-processing is a crucial step prior to continue with other steps required in a ML classification system.
intermediate results, show results of different options
The evolution of the loss, the evolution of the recall and the evolution of the F1 Score. The visualization of interpretability tools (SHAP, Feature Importance...) can be an add on.
n/a

Besides the way the selection of features and hyperparameters affect the final result it would also be useful to see how individual training samples and slight changes in them affect the training process and the overall outcome.

pre-process steps: in order to identify possible outliers or anomalies in the dataset

Hyperparameter/ variable importances-dependencies

The most crucial step would be error analysis, with interactivity to see details for an individual instance and allow some forms of group formation based on confidence. An important property to visualize would be the learned features.

It depends on the task/goal of the analysis. But in general I would like to be in control about what I want to visualize, so the functionalities should be there if I want to use them.

Mainly the basic results. I don't need visualization for analyzing but I need it to communicate about my results to students or collegues. Thus, it's good to have concerns about the conditions of validity but my first goal would be to be able to understand the sociological results.

I would like to see what data points were used for training.

According to my answers it is obvious that ideally I would like to receive feedback on most of the steps. However, if i have to choose one or two, I would say data preparation step is very important - specifically visualising the 'raw' data if possible as that could influence the decision on choosing the appropriate model.

The process of selecting good sampling points and avoiding noisy ones.

Visualize the data transformation in all ML model steps/ and see what is inside the ML black boxs

Training and evaluation/cross-validation steps. Additionally, demonstrating the robustness of the model with regard to noisy and distorted data is also valuable.

validation. statistical improvement over iteration/time. Comparison to non ML methodology/outcome

The decision making part.

The Evaluation step

Additional Comments/Elaboration (8 responses):

I feel bias as an ML researcher, for I might feel more trust in my judgement of an algorithm's performance if I'm acquainted with it, regardless of the domain. E.g., a doctor might want to corroborate some pre-conceptions, while I wouldn't.

The issue here is scalability, some problems can be explored/modelled by users to a lot of detail, but if they have to deal with a massive number of problems (e.g. in time series forecasting 1000s of series/models) then visualisation should also help them to drill down in the cases that are problematic, rather than provide detailed visualisation per series (at least at first!).

Hope this helps!

There is an uncertainty in your questions on what "visualize" means/entails. I've answered the questions as "visualized" (e.g., "Details about the data collection process should be visualized") as "being made aware to" the user. This might be through visual means or through some other interactive modalities.

trust is subjective, hard to generalize

n/a

Hope to hear from the tool if you come up with one (for instance on "quanti" a nice useful mailing list)

I disagree with the use of interaction with the data (and to some extent the model) during training. Such mechanisms would only let the user fish for desired result and would obliterate the statistical significance testing of results.