# DimVis: Interpreting Visual Clusters in Dimensionality Reduction With Explainable Boosting Machine

P. Salmanian[1,3] , A. Chatzimparmpas[2] , A. C. Karaca[3] , and R. M. Martins[1]

[1]Department of Computer Science and Media Technology, Linnaeus University, Sweden
[2]Department of Information and Computing Sciences, Utrecht University, The Netherlands
[3]Department of Computer Engineering, Yildiz Technical University, Turkey
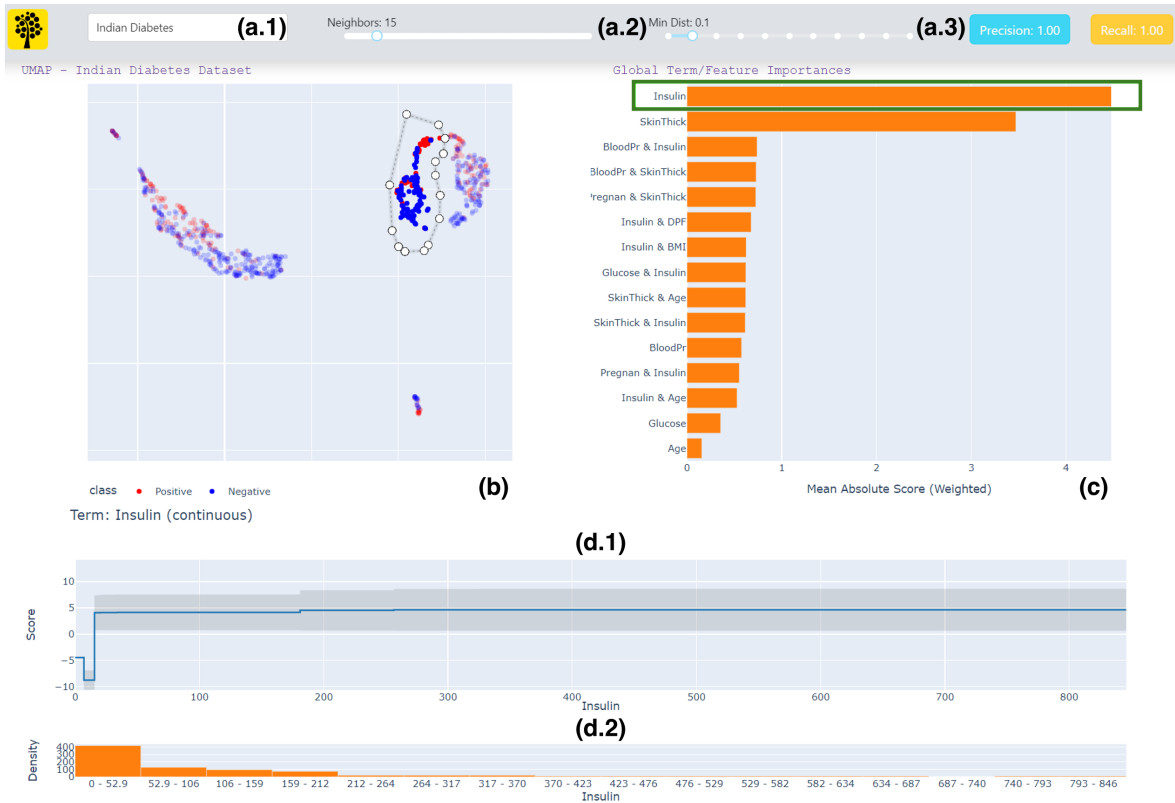
Figure 1: *Components of* DimVis*: a panel for (a.1) choosing a dataset and (a.2) adjusting UMAP hyperparameters, and two performance metrics (a.3) related to the underlying EBM model that supports the visual exploration; (b) a UMAP projection of the dataset, with a selection of points that determine a visual cluster of interest; (c) shows a ranking of features (single and pairs) for the selected visual cluster, based on the EBM model's feature importance; and (d.1) presents the Score of the selection, while (d.2) shows the distribution for one feature.*

Abstract

Dimensionality Reduction (DR) techniques such as t-SNE and UMAP are popular for transforming complex datasets into simpler visual representations. However, while effective in uncovering general dataset patterns, these methods may introduce artifacts and suffer from interpretability issues. This paper presents DimVis, a visualization tool that employs supervised Explainable Boosting Machine (EBM) models (trained on user-selected data of interest) as an interpretation assistant for DR projections. Our tool facilitates high-dimensional data analysis by providing an interpretation of feature relevance in visual clusters through interactive exploration of UMAP projections. Specifically, DimVis uses a contrastive EBM model that is trained in real time to differentiate between the data inside and outside a cluster of interest. Taking advantage of the inherent explainable nature of the EBM, we then use this model to interpret the cluster itself via single and pairwise feature comparisons in a ranking based on the EBM model's feature importance. The applicability and effectiveness of DimVis are demonstrated via a use case and a usage scenario with real-world data. We also discuss the limitations and potential directions for future research.

CCS Concepts

• **Human-centered computing** → Visualization; Visual analytics; • **Machine learning** → Unsupervised learning;

1. Introduction

Dimensionality Reduction (DR) is an important component of the visual exploration of complex and high-dimensional datasets, transforming them into more manageable (but still representative) lower-dimensional forms with the aim to facilitate interpretation and improve computational efficiency while retaining essential information for analysis [SZS*17]. However, the outcomes of modern nonlinear DR techniques such as t-SNE [MH08] and UMAP [MHM18] can be challenging to interpret due to the lack of inherent meaning to the extracted shapes/clusters, suboptimal hyperparameters (e.g., [WVJ16]), and misleading distortions introduced during the process [HFA17, NA18]. To address these challenges, some visualization tools have been developed with the goal of inspecting DR projections interactively to assess the global and local quality of the 2D space [CMJK20, CMJ*20, CKK24] and to produce meaningful interpretations to the extracted shapes and clusters [FKM19, MJE21]. One recent example from Bibal et al. [BCDF21, BDF23] involved training an interpretable supervised Machine Learning (ML) model, the decision tree (DT), and use it to interpret DR layouts. However, DTs suffer from scalability issues, getting quickly deep and unparsable, and may overfit the data to achieve high fidelity. One way to overcome these limitations is by using the Explainable Boosting Machine (EBM) [LCGH13], a tree-based, cyclic gradient boosting generalized additive model, which is a powerful alternative that remains interpretable and scalable even with large-scale, high-dimensional datasets [CLG*15].

We propose a new visualization tool, DimVis, which employs EBM models for unveiling the factors influencing cluster formation and data relationships in nonlinear projections. It enables users to interactively inspect visual patterns of the DR layout (clusters, shapes, etc.) to gain insights about their content by training an EBM model on-the-fly on user-selected data points. The explanation is based on exploring single and pairs of features ranked by EBM's intrinsic feature importance regarding their ability to explain the separation between the selection and the remaining points of the dataset. The results are visualized with intuitive and simple visualizations such as bar charts, line plots, histograms, and heatmaps. We demonstrate DimVis via a use case and a usage scenario with healthcare datasets (also used, for example, in [MQB19] and [CMK20]), showcasing the tool's capability of analyzing high-dimensional datasets and enhancing the interpretability of the resulting 2D space from a DR technique.

2. Related Work

In early research on DR interpretation, static visualizations were generated by considering the layout as a whole, without giving the users the ability and flexibility to specify their own regions of interest [dSRM*15]. Interactive interpretation of selections and groups, as is the case for DimVis, has been explored for example in Probing Projections [SDMT16], where groups of points (obtained either by selection or a clustering algorithm) are compared to each other or to the rest of the dataset via individual histograms for each dimension; the users then must visually determine dimension importance by themselves. This approach provides a strong foundation but may present problems with very high-dimensional datasets, and the interpretations are inherently univariate.

Some examples of recent work that are more related to DimVis are those from Marcilio-Jr et al. [MJE21] and Fujiwara et al. [FKM19]. In both cases, users can define groups of points of interest and interpret them by contrasting their features against outside points. Fujiwara et al. use an adapted version of contrastive PCA [AZBZ18], which in itself is an adapted version of the classic PCA for highlighting the differences between two sets of data points. Although relevant, we argue that being a linear algorithm, interpretations based on adapted versions of PCA will still have limitations when dealing with very high-dimensional and complex groups of points. Marcilio-Jr et al. [MJE21] used Shapley values to generate explanations for each data point based on its distance to the centroids of the specified clusters, which are then visualized with a relatively complex visual abstraction based on merging histograms and connecting them with edges (as lines). In contrast, DimVis uses basic, simple visualizations that are arguably more natural and easily interpreted, and the explanations come directly from an inherently interpretable and robust supervised model instead of being generated with a black-box, model-agnostic method.

DT-SNE [BDF23] is a simpler but *static* approach combining t-SNE with supervised decision trees for interpreting projections with features, and IXVC [BCDF21] supports users in interactively explaining projections with decision trees. However, decision trees, especially visualized as node-link diagrams, can become quickly hard to interpret, and the increase in the number of dataset features may heavily affect their scalability. On the contrary, DimVis uses EBM – which can be as powerful as ensemble learning models – and offers simple and scalable visual representations for exploring the most important features in user-selected data contrastively.

3. DimVis: System Overview and Use Case

DimVis is an open-source, web-based visualization tool [Dim23] which uses the state-of-the-art, supervised "glass-box" EBM model [LCGH13] to interpret visualizations generated with unsupervised DR techniques. We chose the EBM because its results and pre-defined visualizations (which we used) have already been successfully evaluated with real-world domain experts [CLG*15].

After selecting the dataset (shown in Figure 1(a.1)), initially only views indicated in (a.1), (a.2), and (b) are shown. To initiate the exploration and trigger the remaining views, a lasso selection of data points must be made in the UMAP view (b) containing a visual cluster perceived by the user as being of interest. As soon as a selection is made, an EBM model is trained in the background (more details below) and the remaining visualizations are presented. The model's precision and recall are visible (in Figure 1(a.3)), as well as single and pairwise feature importances extracted from the EBM model and visualized in a bar chart sorted in descending order from top to bottom (see Figure 1(c)). On click, users can explore further a feature with a Score line plot (Figure 1(d.1)) and a Density histogram (Figure 1(d.2)), or two features with a Score heatmap.

We further explain the tool by introducing a first use case with the *Breast Cancer Wisconsin (Original)* dataset (entries: 699; dimensions: 9 (graded 1–10); classes: 2) obtained from the UCI ML repository [DG17]. The class labels (*Benign* or *Malignant*) are only used as a final verification for the insights (Figure 2(e)).
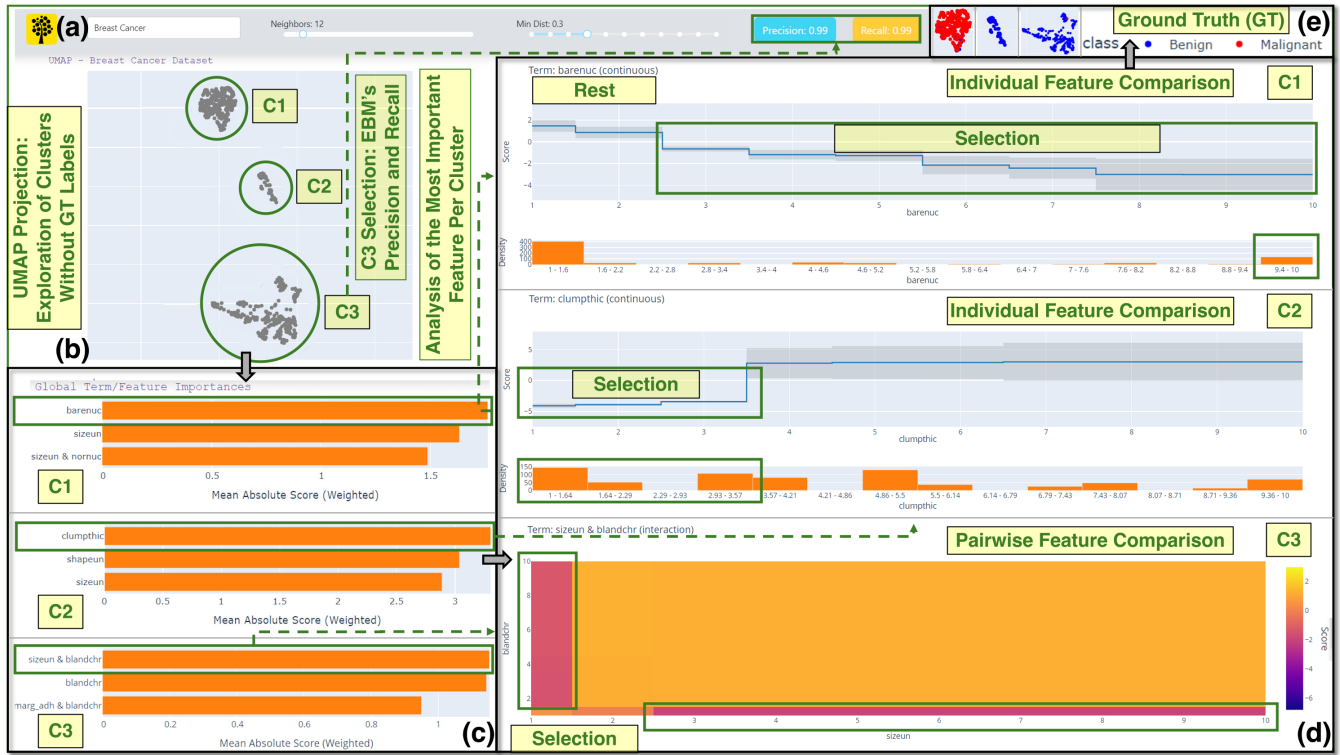
Figure 2: *Exploration of three clusters (C1–C3) produced by UMAP in (b) and local Feature Importances for the top three important features of the EBM model (c), trained on each cluster. (a) shows the dataset, UMAP hyperparameters, and EBM's precision and recall for C3. The most important feature (pair) per cluster is further explained in (d), with the selection mapped with a negative Score and the remaining points a positive one. For a single feature, Density histograms show the distribution of values. (e) reveals the GT labels used solely for verification.*

DIMVIS utilizes the UMAP algorithm [MHM18] and users can interactively adjust UMAP's hyperparameters – "Number of Neighbors" (Neighbors) and "Minimum Distance" (Min Dist) – to explore different projections (see Figure 2(a)). The projection in Figure 2(b) contains three distinct clusters (C1–C3), and no GT labels to simulate a typical unsupervised learning scenario. When a group of points is selected with the lasso functionality, the EBM gets retrained and selected points are marked with class 0 while the remaining points are marked with class 1. Using the inherent interpretability features of the EBM model, we rank the main features (single and pairs) that contribute to the separation of the cluster according to the trained model, using the Mean Absolute Score (Weighted). As shown in Figure 2(c), the most important single and pair of features are: the number of bare nuclei for C1 (BARENUC, ≈1.7), clump thickness for C2 (CLUMPTHIC, ≈3.2), and the combination of uniformity of cell size and bland chromatin for C3 (SIZEUN & BLANDCHR, ≈1.2). That can be further explored in relation to the actual values and the distribution of the data entries (see below). More algorithmic details regarding EBM can also be found in Lou et al. [LCGH13].

When a user clicks on a single feature in the bar chart (Figure 2(c), C1), a line plot and a histogram appear, displaying the impact of that specific feature (BARENUC, in this case) on the model's predictions. As illustrated in Figure 2(d), C1 (top), the line plot shows the *Score*, i.e., the contribution of each feature's value in-side (*negative* values) or outside (*positive* values) of the selection. It also shows the *quantified uncertainty*, that is, the variance in the final predictions, encoded as gray bands. For C2, for example, we can see that values of CLUMPTHIC up to 3.5 are more related to the points inside the selected cluster, with high certainty. The rest of the values are more related to the points outside the cluster, although with less certainty. The histogram depicts the *Density*, which is a simple distribution of the values for a specific feature across all points globally. For C1, the feature BARENUC has a clear separation between low and high values that seems to be related to the cluster formation, while for C2, CLUMPTHIC does not seem to show any clear pattern for points inside or outside the cluster (which shows why it is important to consider this view together with the *Score* line plot). Additionally, DIMVIS allows users to explore pairs of features, as shown in Figure 2(d). The heatmap uses color intensity to represent the level of interaction (i.e., from purple for *negative Score* to yellow for *positive Score*). In Figure 2(d), C3, we see that points where one of the two features is low while the other is high are more related to the cluster (see green boxes); when both features are low, or both are high, then the *Score* is higher, indicating a stronger relation to the points outside the cluster.

In the next section we explore different aspects of the user experience of DIMVIS by simulating a usage scenario in the domain of healthcare, one of the important application areas of DR.
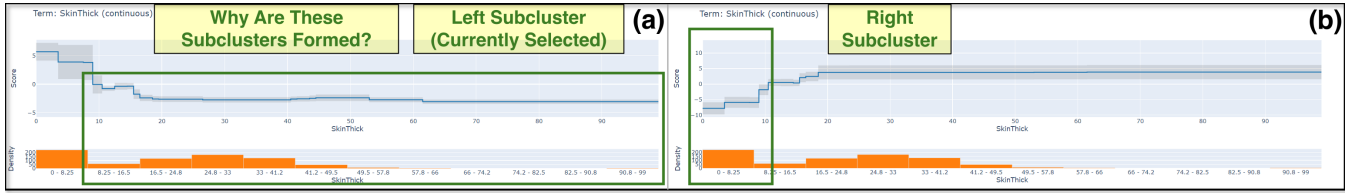
Figure 3: *The individual feature importance of SkinThick explains why the user-selected left and right subclusters in (a) and (b), respectively, were formed. The left subcluster (negative Score) contains points mostly with moderate values for SkinThick (Density) and low uncertainty (see the tight, gray band), while the right subcluster includes points with low values for SkinThick and relatively moderate uncertainty.*

4. Usage Scenario: Interpreting Clusters in a Projection

For this hypothetical scenario, we follow a physician specializing in diabetes (which we will call Sara). She recently learned about UMAP but, while enthusiastic, she remains doubtful about the reliability of such projections. Hence, she turns to DIMVIS for an in-depth analysis of UMAP projections. Sara uploads to DIMVIS (Figure 1(a.1)) the *Pima Indian diabetes* dataset (entries: 768; dimensions: 8; classes: 2) from the UCI ML repository [DG17].

Without any selection, DIMVIS trains a supervised EBM model (precision: 0.81; recall: 0.82) and visualizes the Global Feature Importances from the most to the least important: Glucose, BMI, Age, DPF, Insulin, Pregnan, SkinThick, and BloodPr (not shown due to space limits). Now that she has gained an overview, Sara wants to investigate why a cluster in Figure 1(b) is split into left and right subclusters. By using a lasso to choose the left subcluster (cf. Figure 1(b)), the tool fits the EBM model (precision and recall: 1.00; see Figure 1(a.3)) to this selection of points (against the rest) and recomputes feature importance locally. Insulin and SkinThick are the two most important dimensions and require further examination (see Figure 1(c)). When selecting the right subcluster, the feature ranking produces the opposite result (SkinThick first and then Insulin). She acknowledges that these two dimensions are by far the most important for the formation of this entire cluster.

Sara clicks on the first feature, which is *Insulin* (Figure 1(c), green box), and the Score and Density graphs appear in Figure 1(d.1 and d.2). In Figure 1(d.1), negative Score, she notices that the values of the selected points for Insulin are very low with most entries in the first bucket [0-52.9] (Figure 1(d.2)). This feature follows a linear trend with high uncertainty (see the wide, gray band) for the rest of the data points, as shown in Figure 1(d.1), positive Score. However, the same effect is observable for the right subcluster (omitted due to space limits). She decides to continue with the exploration of the second important feature, which is SkinThick. She checks the individual feature importance of SkinThick for the left and right subclusters, as presented in Figure 3, (a) and (b), respectively. She understands that the former contains mostly points with moderate values (cf. Density graph) and low uncertainty for this dimension (tight gray bar in Score graph; green box), while the latter demonstrates most points concentrated in values [0-8.25] with somewhat moderate uncertainty (see the green box in Figure 3(b)). Notably, Sara identified the opposite effect for SkinThick in the right subcluster compared to what she had seen in the inspection of the left subcluster. She concludes, then, that the common trait for this entire cluster is Insulin, while the reason behind its division into two subclusters is the different values in SkinThick.

5. Discussion

Design choices. Although in its current version DIMVIS uses UMAP and EBM, the architecture of the tool allows it to be model-agnostic and flexible. The choices of EBM and UMAP were made since they are both state-of-the-art methods that can scale to large and high-dimensional datasets, but the flexible design of the work-flow ensures its potential adaptability to emerging, future algorithms. Also, one of the strengths of EBM is that it can be interpreted with simple visualizations of single or pairs of features (as opposed to, for example, visualizing decision trees as non-space-filling node-link diagrams). The visualizations used for DIMVIS are inherited from the EBM model (pre-evaluated, for example, with real-world medical data [CLG*15]) and do not represent a novelty of this paper. As such, a future direction could be to enhance the presentation beyond what is currently used. However, more complex visualizations may have a steeper learning curve and suffer in terms of scalability, effectiveness, and ease of use.

Limitations and future work. One important future work is an objective comparison to other similar tools and techniques so as to determine concretely how the detected explanations (e.g., features and their ranks) differ from other possible methods for generating similar explanations. This is one of our short-term plans for future work. We also intend to refine the user experience based on feedback from domain experts (as in Section 4) and perform a formal user study to validate our findings further and gather more feedback on possible improvements. Different possible visualization techniques (with variable degrees of complexity) will be tested with users to determine whether it is worth it to move beyond the current ones and how to do it. We intend to investigate ways to improve the computational efficiency of the underlying model by, for example, using GPU-based implementations of both the DR and the EBM algorithms. Finally, DIMVIS can currently only compare one data selection against the remaining projection, making a direct comparison between two subclusters a somewhat inefficient task. We intend to test the ability to select and compare two specific data subsets directly to streamline this process.

6. Conclusion

In this paper, we presented DIMVIS, a visualization tool using a supervised EBM model to interpret UMAP projections interactively. Our tool helps users with the in-depth exploration of cluster formations and their analysis based on the impact of single and pairwise features on the data. The applicability and usefulness of DIMVIS were tested with two use cases that showcase diverse settings for understanding real-world, high-dimensional data.

References

[AZBZ18] ABID A., ZHANG M. J., BAGARIA V. K., ZOU J.: Exploring patterns enriched in a dataset with contrastive principal component analysis. *Nature Communications 9* (2018), 2134. `doi:10.1038/s41467-018-04608-8`. 2

[BCDF21] BIBAL A., CLARINVAL A., DUMAS B., FRÉNAY B.: IXVC: An interactive pipeline for explaining visual clusters in dimensionality reduction visualizations with decision trees. *Array 11* (2021), 100080. `doi:10.1016/j.array.2021.100080`. 2

[BDF23] BIBAL A., DELCHEVALERIE V., FRÉNAY B.: DT-SNE: t-SNE discrete visualizations as decision tree structures. *Neurocomputing 529*, C (2023), 101–112. `doi:10.1016/j.neucom.2023.01.073`. 2

[CKK24] CHATZIMPARMPAS A., KUCHER K., KERREN A.: Visualization for trust in machine learning revisited: The state of the field in 2023. *IEEE Computer Graphics and Applications* (2024), 1–14. to appear. `doi:10.1109/MCG.2024.3360881`. 2

[CLG*15] CARUANA R., LOU Y., GEHRKE J., KOCH P., STURM M., ELHADAD N.: Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2015), ACM, pp. 1721–1730. `doi:10.1145/2783258.2788613`. 2, 4

[CMJ*20] CHATZIMPARMPAS A., MARTINS R. M., JUSUFI I., KUCHER K., ROSSI F., KERREN A.: The state of the art in enhancing trust in machine learning models with the use of visualizations. *Computer Graphics Forum* (2020). `doi:10.1111/cgf.14034`. 2

[CMJK20] CHATZIMPARMPAS A., MARTINS R. M., JUSUFI I., KERREN A.: A survey of surveys on the use of visualization for interpreting machine learning models. *Information Visualization 19*, 3 (2020), 207–233. `doi:10.1177/1473871620904671`. 2

[CMK20] CHATZIMPARMPAS A., MARTINS R. M., KERREN A.: t-viSNE: Interactive assessment and interpretation of t-SNE projections. *IEEE Transactions on Visualization and Computer Graphics 26*, 8 (2020), 2696–2714. `doi:10.1109/TVCG.2020.2986996`. 2

[DG17] DUA D., GRAFF C.: UCI Machine Learning Repository, 2017. URL: `https://archive.ics.uci.edu/ml`. 2, 4

[Dim23] DimVis Code, 2023. Accessed April 3, 2024. URL: `https://github.com/parisa-salmanian/DimVis`. 2

[dSRM*15] DA SILVA R., RAUBER P., MARTINS R., MINGHIM R., TELEA A.: Attribute-based visual explanation of multidimensional projections. In *Proceedings of the EuroVA* (2015), pp. 134–139. 2

[FKM19] FUJIWARA T., KWON O.-H., MA K.-L.: Supporting analysis of dimensionality reduction results with contrastive learning. *IEEE Transactions on Visualization and Computer Graphics 26*, 1 (2019), 45–55. `doi:10.1109/TVCG.2019.2934251`. 2

[HFA17] HEULOT N., FEKETE J.-D., AUPETIT M.: Visualizing dimensionality reduction artifacts: An evaluation. *ArXiv e-prints* (2017). `arXiv:1705.05283`. 2

[LCGH13] LOU Y., CARUANA R., GEHRKE J., HOOKER G.: Accurate intelligible models with pairwise interactions. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2013), KDD '13, ACM, pp. 623–631. `doi:10.1145/2487575.2487579`. 2, 3

[MH08] MAATEN L. V. D., HINTON G.: Visualizing data using t-SNE. *Journal of Machine Learning Research 9*, Nov (2008), 2579–2605. URL: `https://jmlr.org/papers/v9/vandermaaten08a.html`. 2

[MHM18] MCINNES L., HEALY J., MELVILLE J.: UMAP: Uniform Manifold Approximation and Projection for dimension reduction. *ArXiv e-prints* (2018). `arXiv:1802.03426`. 2, 3

[MJE21] MARCÍLIO-JR W. E., ELER D. M.: Explaining dimensionality reduction results using shapley values. *Expert Systems with Applications 178* (2021), 115020. `doi:10.1016/j.eswa.2021.115020`. 2

[MQB19] MING Y., QU H., BERTINI E.: RuleMatrix: Visualizing and Understanding Classifiers with Rules. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2019), 342–352. `doi:10.1109/TVCG.2018.2864812`. 2

[NA18] NONATO L. G., AUPETIT M.: Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment. *IEEE Transactions on Visualization and Computer Graphics 25*, 8 (2018), 2650–2673. `doi:10.1109/TVCG.2018.2846735`. 2

[SDMT16] STAHNKE J., DÖRK M., MÜLLER B., THOM A.: Probing projections: Interaction techniques for interpreting arrangements and errors of dimensionality reductions. *IEEE Transactions on Visualization and Computer Graphics 22* (2016), 629–638. `doi:10.1109/TVCG.2015.2467717`. 2

[SZS*17] SACHA D., ZHANG L., SEDLMAIR M., LEE J. A., PELTONEN J., WEISKOPF D., NORTH S. C., KEIM D. A.: Visual interaction with dimensionality reduction: A structured literature analysis. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2017), 241–250. `doi:10.1109/TVCG.2016.2598495`. 2

[WVJ16] WATTENBERG M., VIÉGAS F., JOHNSON I.: How to use t-SNE effectively. *Distill* (2016). `doi:10.23915/distill.00002`. 2