Computational Statistics

Exploring Local Explanations of Nonlinear Models Using Animated Linear Projections

--Manuscript Draft--

Manuscript Number:	COST-D-23-00336R1
Full Title:	Exploring Local Explanations of Nonlinear Models Using Animated Linear Projections
Article Type:	Software Article
Keywords:	explainable artificial intelligence; nonlinear model interpretability; visual analytics; local explanations; grand tour; radial tour
Manuscript Classifications:	1.00600: Cluster Analysis; 1.02490: Machine Learning; 1.02830: Multidimensional Data Analysis; 1.02840: Multidimensional Visualization; 1.03040: Nonlinear Regression; 1.04290: Software: R; 1.04520: Statistical Graphics

Corresponding Author:	Nicholas Spyrison Monash University AUSTRALIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Monash University
Corresponding Author's Secondary Institution:	
First Author:	Nicholas Spyrison
First Author Secondary Information:	
Order of Authors:	Nicholas Spyrison
	Dianne Cook
	Przemysław Biecek
Order of Authors Secondary Information:	

Funding Information:	Australian Government	Dr. Nicholas Spyrison

Abstract:	The increased predictive power of machine learning models comes at the cost of increased complexity and loss of interpretability, particularly in comparison to parametric statistical models. This trade-off has led to the emergence of eXplainable AI (XAI) which provides methods, such as local explanations (LEs) and local variable attributions (LVAs), to shed light on how a model use predictors to arrive at a prediction. These provide a point estimate of the linear variable importance in the vicinity of a single observation. However, LVAs tend not to effectively handle association between predictors. To understand how the interaction between predictors affects the variable importance estimate, we can convert LVAs into linear projections and use the radial tour. This is also useful for learning how a model has made a mistake, or the effect of outliers, or the clustering of observations. The approach is illustrated with examples from categorical (penguin species, chocolate types) and quantitative (soccer/football salaries, house prices) response models. The methods are implemented in the R package cheem, available on CRAN.

Response to review

Cheem Exploring Local Explanations of Nonlinear Models Using Animated Linear Projections
November 10, 2023

Thank you to the editor and reviewers for your helpful feedback on our paper. We have revisited the manuscript and to address the reviewer's comments:

We respond the reviewer comments inline in blue, and large changes in the paper are also marked in blue.

Changes made also reflect a change in the R package, with the cheem v0.4.0 is now hosted on CRAN as of 11/08/2023.

Thank you,
Nicholas Spyrison,
Corresponding author

COMMENTS FOR THE AUTHOR:

The authors should update their references using the Computational Statistics citation style.

We now follow the guidelines provided at the journal web site and use the Springer Nature template as suggested.

Reviewer #1: This paper proposes an interactive visualization tool and its implementation as an R package to intuitively understand LVA in non-linear models.

General Comments

You have not clearly stated the usefulness of the radial tour; please explain the advantage of using the SHAP values of PI and CI as the contribution of 1D projection compared to drawing a static density plot or scatterplot. Is it possible to correctly evaluate the sensitivity of variables by 1d projection for nonlinear models? Also, in the Case Studies in Section 5, please provide an explanation in relation to the Radial tour operation.

Section 3 introduced the radial tour and provided a simple explanation of how it is generally used for understanding variable importance. Section 4.2 connects this to the use of SHAP

values for explaining the prediction for an individual observation. We have added a paragraph with the general thinking.

SHAP values are effectively locally linear estimates of variable importance for a prediction. Thus, yes, in the local neighbourhood of an observation, testing the sensitivity of variables for the prediction can be done using 1D projections. We have added a sentence to this effect.

We have also added some additional wording in Section 5 case studies.

The results are very different between the examples provided in the shiny application in the R cheem package and the examples presented in the paper. Please make sure to use random number seeds to achieve similar results. In particular, in the Case Studies of classification, there is no misclassification in the examples in the cheem package, and the misclassification is not indicated by the red circle as shown in the paper.

We have revisited the attached models and objects and verified that classification models misclassify some observations for the purposes of discussion.

The panel that is treeshap in the global view in the paper is attribution in the cheem package. Please unify either one.

We have clarified what the panels depict in the application and figures 3 and 4. This will meaningfully vary from other figures In this paper. As we use specific models and attributions to illustrate, while the package allows for the general case.

In the global view in the paper, the model panel has jitter processing, but the cheem package does not.

Classification models do use jitter to avoid occlusion of the observations. While regression models have more continuous distributions. This has been clarified in the panel titles.

Discussion is just a Conclusion. For example, please supplement any discussion on the following.

- Abstract and Introduction are described in the context of XAI, but the tools proposed in this paper are not likely to be able to handle large data sets.

Done, changes are highlighted in blue.

Concerning the convenience of the package: the application by the cheem package outputs little information (e.g., axis information), which makes it difficult to use intuitively.

We have revisited the panel strips to better describe what is being viewed.

Individual Comments

(p.7 l.18) Figures 3b, 4a -> 3c, 4b?

You are correct, this has been fixed.

(p.8 l.7) Is the SHAP value of the CI taken into account in the radial tour?

In the explanations of the radial tour we have more clearly stated that it starts from the PI SHAP values. When changing the projection values one could make it closer to those of the CI, which is done in some examples.

(p.8 l.8) "The PI is ... global view" : The shiny app in the cheem package does not support interactive selection by clicking in the global view. (cheem version 0.3.0 on Windows 11 / Google Chrome)

We have clarified the text from selection of the PI to click and drag highlighting discussed above and below this point. We have confirmed that plotly interaction is listed as an import. And validated the interaction (cheem version 0.3.0 on Windows 10 / Google Chrome).

(p.10 Fig4) Residual plot is not displayed in cheem version 0.3.0 on Windows 11 / Google Chrome.

We have confirmed that residual plots do display for regression models, while classification models give more of a visual confusion matrix display. (cheem version 0.3.0 on Windows 10 / Google Chrome).

(p.11 l.5) Figure 3a & b -> 4a & b?

We have corrected the manuscript, thank you.

(p.11 l.7) Figure 3d -> 4d?

We have corrected the manuscript, thank you.

(p.21 l.10) https://nspyrison.github.io/cheem/ is not available.

The packagedown site has been recovered, updated version has been submitted to CRAN.

(references) There are some characters like "???".

Fixed

Reviewer #2: This is a very interesting paper.
I was able to run your software using publicly available packages.
Unfortunately, I could not play videos shown at the following URLs.
https://vimeo.com/666431172
https://vimeo.com/666431143
https://vimeo.com/666431148
https://vimeo.com/666431163
https://vimeo.com/666431134

It would be even easier to understand if videos could be played back, so I hope this will be possible.

We cannot address this unless the error messages are provided. The videos at each of the addresses are playable by us and several colleagues who also checked. We have validated that these url's are pointing to the correct content, videos are set to public and content is marked for all audiences.

I think the visualization you propose is very clear.
However, I have the impression that it is difficult to understand, so I expect detailed explanations.
For example, detailed explanations of Figure 3 e and Figure 4 e would be helpful.
I also think that explanations of the data in Figure 3 and Figure 4 would make it even easier to understand.

See the response to reviewer 1. The new text gives more details that should make it easier to understand.

For small details, I think it is necessary to check page 5, line 38 and page 11, line 29.

Page 5, line 38, discusses the R packages related to LVA. These packages have been validated to be available on CRAN. treeshap has been updated to reflect it's availability on CRAN. Text and references have also been updated for accuracy version number at time of clarification 10/24/2023.

Page 11, line 29 discusses the hyperparameter values used in the random forest models. This sentence has been parsed into two for readability, regression and classification respectively. The values have been validated to still be accurate to the examples used.

```
This is pdfTeX, Version 3.141592653-2.6-1.40.24 (TeX Live 2022)
(preloaded format=pdflatex 2023.3.8)  10 NOV 2023 16:31
entering extended mode
 restricted \write18 enabled.
 %&-line parsing enabled.
**sn-article.tex
(./sn-article.tex
LaTeX2e <2022-11-01> patch level 1
L3 programming layer <2023-02-22> (c:/TeXLive/texmf-
local/tex/latex/aries/sn-jn
l.cls
Document Class: sn-jnl 2019/11/18 v0.1: An authoring template for
Springer Jour
nal articles
(c:/TeXLive/2022/texmf-dist/tex/latex/base/article.cls
Document Class: article 2022/07/02 v1.4n Standard LaTeX document class
(c:/TeXLive/2022/texmf-dist/tex/latex/base/fleqn.clo
File: fleqn.clo 2016/12/29 v1.2b Standard LaTeX option (flush left
equations)
\mathindent=\skip48
Applying: [2015/01/01] Make \[ robust on input line 50.
LaTeX Info: Redefining \[ on input line 51.
Already applied: [0000/00/00] Make \[ robust on input line 62.
Applying: [2015/01/01] Make \] robust on input line 74.
LaTeX Info: Redefining \] on input line 75.
Already applied: [0000/00/00] Make \] robust on input line 83.
) (c:/TeXLive/2022/texmf-dist/tex/latex/base/size10.clo
File: size10.clo 2022/07/02 v1.4n Standard LaTeX file (size option)
)
\c@part=\count185
\c@section=\count186
\c@subsection=\count187
\c@subsubsection=\count188
\c@paragraph=\count189
\c@subparagraph=\count190
\c@figure=\count191
\c@table=\count192
\abovecaptionskip=\skip49
\belowcaptionskip=\skip50
\bibindent=\dimen140
)
LaTeX Info: Redefining \rm on input line 144.
LaTeX Info: Redefining \sf on input line 145.
LaTeX Info: Redefining \tt on input line 146.
LaTeX Info: Redefining \bf on input line 147.
LaTeX Info: Redefining \it on input line 148.
LaTeX Info: Redefining \sl on input line 149.
LaTeX Info: Redefining \sc on input line 150.
LaTeX Info: Redefining \cal on input line 151.
LaTeX Info: Redefining \mit on input line 152.
LaTeX Info: Redefining \textsubscript on input line 210.
\columnhsize=\skip51
(c:/TeXLive/2022/texmf-dist/tex/latex/geometry/geometry.sty
Package: geometry 2020/01/02 v5.9 Page Geometry
```

```
(c:/TeXLive/2022/texmf-dist/tex/latex/graphics/keyval.sty
Package: keyval 2022/05/29 v1.15 key=value parser (DPC)
\KV@toks@=\toks16
) (c:/TeXLive/2022/texmf-dist/tex/generic/iftex/ifvtex.sty
Package: ifvtex 2019/10/25 v1.7 ifvtex legacy package. Use iftex instead.
(c:/TeXLive/2022/texmf-dist/tex/generic/iftex/iftex.sty
Package: iftex 2022/02/03 v1.0f TeX engine tests
))
\Gm@cnth=\count193
\Gm@cntv=\count194
\c@Gm@tempcnt=\count195
\Gm@bindingoffset=\dimen141
\Gm@wd@mp=\dimen142
\Gm@odd@mp=\dimen143
\Gm@even@mp=\dimen144
\Gm@layoutwidth=\dimen145
\Gm@layoutheight=\dimen146
\Gm@layouthoffset=\dimen147
\Gm@layoutvoffset=\dimen148
\Gm@dimlist=\toks17
)
\artcatbox=\box51
\aucount=\count196
\corraucount=\count197
\punctcount=\count198
\emailcnt=\count199
\c@affn=\count266
\addcount=\count267
\PacsCount=\count268
\PacsTmpCnt=\count269
\FMremarkdim=\dimen149
\fmremarkbox=\box52
(c:/TeXLive/2022/texmf-dist/tex/latex/sttools/cuted.sty
Package: cuted 2021/10/04 v2.0 Mixing onecolumn and twocolumn modes
\At@ViperColsBreak=\toks18
\preCutedStrip=\toks19
\postCutedStrip=\toks20
\cuted@@tempbox@a=\box53
\cuted@@tempbox@c=\box54
\cuted@@tempbox@var=\box55
\hold@viper=\box56
\@viper=\box57
\cuted@@varbox@a=\box58
\cuted@@varbox@c=\box59
\cuted@@tempdim@spread=\dimen150
\cuted@@tempdim@a=\dimen151
\cuted@@tempdim@b=\dimen152
\ht@hold@viper=\dimen153
\ht@viper=\dimen154
\var@@pagediscards@ht=\dimen155
\stripsep=\skip52
\viper@penalty=\count270
)
\firstpagehtcheck=\dimen156
```

```
\labelwidthi=\dimen157
\labelwidthii=\dimen158
\labelwidthiii=\dimen159
\labelwidthiv=\dimen160
\figwidth=\dimen161
\figheight=\dimen162
\sidecapwidth=\dimen163
\wrapcapline=\dimen164
\totalwrapline=\dimen165
\wraptotline=\dimen166
\figurebox=\box60
\wrapfigcapbox=\box61
\figcapbox=\box62
\capbox=\box63
\headwidthskip=\skip53
\tabcapbox=\box64
\temptbox=\box65
\tempdime=\dimen167
\tabhtdime=\dimen168
(c:/TeXLive/2022/texmf-dist/tex/latex/graphics/rotating.sty
Package: rotating 2016/08/11 v2.16d rotated objects in LaTeX
(c:/TeXLive/2022/texmf-dist/tex/latex/graphics/graphicx.sty
Package: graphicx 2021/09/16 v1.2d Enhanced LaTeX Graphics (DPC,SPQR)
(c:/TeXLive/2022/texmf-dist/tex/latex/graphics/graphics.sty
Package: graphics 2022/03/10 v1.4e Standard LaTeX Graphics (DPC,SPQR)
(c:/TeXLive/2022/texmf-dist/tex/latex/graphics/trig.sty
Package: trig 2021/08/11 v1.11 sin cos tan (DPC)
) (c:/TeXLive/2022/texmf-dist/tex/latex/graphics-cfg/graphics.cfg
File: graphics.cfg 2016/06/04 v1.11 sample graphics configuration
)
Package graphics Info: Driver file: pdftex.def on input line 107.
(c:/TeXLive/2022/texmf-dist/tex/latex/graphics-def/pdftex.def
File: pdftex.def 2022/09/22 v1.2b Graphics/color driver for pdftex
))
\Gin@req@height=\dimen169
\Gin@req@width=\dimen170
) (c:/TeXLive/2022/texmf-dist/tex/latex/base/ifthen.sty
Package: ifthen 2022/04/13 v1.1d Standard LaTeX ifthen package (DPC)
)
\c@r@tfl@t=\count271
\rotFPtop=\skip54
\rotFPbot=\skip55
\rot@float@box=\box66
\rot@mess@toks=\toks21
) (c:/TeXLive/2022/texmf-dist/tex/latex/threeparttable/threeparttable.sty
Package: threeparttable 2003/06/13  v 3.0
\@tempboxb=\box67
) (c:/TeXLive/2022/texmf-dist/tex/latex/appendix/appendix.sty
Package: appendix 2020/02/08 v1.2c extra appendix facilities
\c@@pps=\count272
\c@@ppsavesec=\count273
\c@@ppsaveapp=\count274
) (c:/TeXLive/2022/texmf-dist/tex/latex/hyperref/hyperref.sty
Package: hyperref 2023-02-07 v7.00v Hypertext links for LaTeX
```

```
(c:/TeXLive/2022/texmf-dist/tex/generic/ltxcmds/ltxcmds.sty
Package: ltxcmds 2020-05-10 v1.25 LaTeX kernel commands for general use
(HO)
) (c:/TeXLive/2022/texmf-dist/tex/generic/pdftexcmds/pdftexcmds.sty
Package: pdftexcmds 2020-06-27 v0.33 Utility functions of pdfTeX for
LuaTeX (HO
)
(c:/TeXLive/2022/texmf-dist/tex/generic/infwarerr/infwarerr.sty
Package: infwarerr 2019/12/03 v1.5 Providing info/warning/error messages
(HO)
)
Package pdftexcmds Info: \pdf@primitive is available.
Package pdftexcmds Info: \pdf@ifprimitive is available.
Package pdftexcmds Info: \pdfdraftmode found.
) (c:/TeXLive/2022/texmf-dist/tex/latex/kvsetkeys/kvsetkeys.sty
Package: kvsetkeys 2022-10-05 v1.19 Key value parser (HO)
) (c:/TeXLive/2022/texmf-dist/tex/generic/kvdefinekeys/kvdefinekeys.sty
Package: kvdefinekeys 2019-12-19 v1.6 Define keys (HO)
) (c:/TeXLive/2022/texmf-dist/tex/generic/pdfescape/pdfescape.sty
Package: pdfescape 2019/12/09 v1.15 Implements pdfTeX's escape features
(HO)
) (c:/TeXLive/2022/texmf-dist/tex/latex/hycolor/hycolor.sty
Package: hycolor 2020-01-27 v1.10 Color options for hyperref/bookmark
(HO)
) (c:/TeXLive/2022/texmf-dist/tex/latex/letltxmacro/letltxmacro.sty
Package: letltxmacro 2019/12/03 v1.6 Let assignment for LaTeX macros (HO)
) (c:/TeXLive/2022/texmf-dist/tex/latex/auxhook/auxhook.sty
Package: auxhook 2019-12-17 v1.6 Hooks for auxiliary files (HO)
) (c:/TeXLive/2022/texmf-dist/tex/latex/hyperref/nameref.sty
Package: nameref 2022-05-17 v2.50 Cross-referencing by name of section
(c:/TeXLive/2022/texmf-dist/tex/latex/refcount/refcount.sty
Package: refcount 2019/12/15 v3.6 Data extraction from label references
(HO)
) (c:/TeXLive/2022/texmf-
dist/tex/generic/gettitlestring/gettitlestring.sty
Package: gettitlestring 2019/12/15 v1.6 Cleanup title references (HO)
(c:/TeXLive/2022/texmf-dist/tex/latex/kvoptions/kvoptions.sty
Package: kvoptions 2022-06-15 v3.15 Key value format for package options
(HO)
))
\c@section@level=\count275
)
\@linkdim=\dimen171
\Hy@linkcounter=\count276
\Hy@pagecounter=\count277
(c:/TeXLive/2022/texmf-dist/tex/latex/hyperref/pd1enc.def
File: pd1enc.def 2023-02-07 v7.00v Hyperref: PDFDocEncoding definition
(HO)
Now handling font encoding PD1 ...
... no UTF-8 mapping file for font encoding PD1
) (c:/TeXLive/2022/texmf-dist/tex/generic/intcalc/intcalc.sty
Package: intcalc 2019/12/15 v1.3 Expandable calculations with integers
(HO)
) (c:/TeXLive/2022/texmf-dist/tex/generic/etexcmds/etexcmds.sty
```

```
Package: etexcmds 2019/12/15 v1.7 Avoid name clashes with e-TeX commands
(HO)
)
\Hy@SavedSpaceFactor=\count278
(c:/TeXLive/2022/texmf-dist/tex/latex/hyperref/puenc.def
File: puenc.def 2023-02-07 v7.00v Hyperref: PDF Unicode definition (HO)
Now handling font encoding PU ...
... no UTF-8 mapping file for font encoding PU
)
Package hyperref Info: Hyper figures OFF on input line 4177.
Package hyperref Info: Link nesting OFF on input line 4182.
Package hyperref Info: Hyper index ON on input line 4185.
Package hyperref Info: Plain pages OFF on input line 4192.
Package hyperref Info: Backreferencing OFF on input line 4197.
Package hyperref Info: Implicit mode ON; LaTeX internals redefined.
Package hyperref Info: Bookmarks ON on input line 4425.
\c@Hy@tempcnt=\count279
(c:/TeXLive/2022/texmf-dist/tex/latex/url/url.sty
\Urlmuskip=\muskip16
Package: url 2013/09/16  ver 3.4  Verb mode for urls, etc.
)
LaTeX Info: Redefining \url on input line 4763.
\XeTeXLinkMargin=\dimen172
(c:/TeXLive/2022/texmf-dist/tex/generic/bitset/bitset.sty
Package: bitset 2019/12/09 v1.3 Handle bit-vector datatype (HO)
(c:/TeXLive/2022/texmf-dist/tex/generic/bigintcalc/bigintcalc.sty
Package: bigintcalc 2019/12/15 v1.5 Expandable calculations on big
integers (HO
)
))
\Fld@menulength=\count280
\Field@Width=\dimen173
\Fld@charsize=\dimen174
Package hyperref Info: Hyper figures OFF on input line 6042.
Package hyperref Info: Link nesting OFF on input line 6047.
Package hyperref Info: Hyper index ON on input line 6050.
Package hyperref Info: backreferencing OFF on input line 6057.
Package hyperref Info: Link coloring OFF on input line 6062.
Package hyperref Info: Link coloring with OCG OFF on input line 6067.
Package hyperref Info: PDF/A mode OFF on input line 6072.
(c:/TeXLive/2022/texmf-dist/tex/latex/base/atbegshi-ltx.sty
Package: atbegshi-ltx 2021/01/10 v1.0c Emulation of the original atbegshi
package with kernel methods
)
\Hy@abspage=\count281
\c@Item=\count282
\c@Hfootnote=\count283
)
Package hyperref Info: Driver (autodetected): hpdftex.
(c:/TeXLive/2022/texmf-dist/tex/latex/hyperref/hpdftex.def
File: hpdftex.def 2023-02-07 v7.00v Hyperref driver for pdfTeX
(c:/TeXLive/2022/texmf-dist/tex/latex/base/atveryend-ltx.sty
Package: atveryend-ltx 2020/08/19 v1.0a Emulation of the original
atveryend pac
```

```
kage
with kernel methods
)
\Fld@listcount=\count284
\c@bookmark@seq@number=\count285
(c:/TeXLive/2022/texmf-dist/tex/latex/rerunfilecheck/rerunfilecheck.sty
Package: rerunfilecheck 2022-07-10 v1.10 Rerun checks for auxiliary files
(HO)
(c:/TeXLive/2022/texmf-dist/tex/generic/uniquecounter/uniquecounter.sty
Package: uniquecounter 2019/12/15 v1.4 Provide unlimited unique counter
(HO)
)
Package uniquecounter Info: New unique counter `rerunfilecheck' on input
line 2
85.
)
\Hy@SectionHShift=\skip56
) (c:/TeXLive/2022/texmf-dist/tex/latex/breakurl/breakurl.sty
Package: breakurl 2013/04/10 v1.40 Breakable hyperref URLs
(c:/TeXLive/2022/texmf-dist/tex/latex/xkeyval/xkeyval.sty
Package: xkeyval 2022/06/16 v2.9 package option processing (HA)
(c:/TeXLive/2022/texmf-dist/tex/generic/xkeyval/xkeyval.tex
(c:/TeXLive/2022/te
xmf-dist/tex/generic/xkeyval/xkvutils.tex
\XKV@toks=\toks22
\XKV@tempa@toks=\toks23
)
\XKV@depth=\count286
File: xkeyval.tex 2014/12/03 v2.7a key=value parser (HA)
)) (c:/TeXLive/2022/texmf-dist/tex/generic/iftex/ifpdf.sty
Package: ifpdf 2019/10/25 v3.4 ifpdf legacy package. Use iftex instead.
)

Package breakurl Warning: You are using breakurl while processing via
pdflatex.

(breakurl)                \burl will be just a synonym of \url.
(breakurl)                 on input line 48.

)
Package hyperref Info: Option `colorlinks' set `true' on input line 1483.
Package hyperref Info: Option `breaklinks' set `true' on input line 1483.
Package hyperref Info: Option `plainpages' set `false' on input line
1483.
Package hyperref Info: Option `bookmarksopen' set `true' on input line
1483.
Package hyperref Info: Option `bookmarksnumbered' set `false' on input
line 148
3.
(c:/TeXLive/2022/texmf-dist/tex/latex/wrapfig/wrapfig.sty
\wrapoverhang=\dimen175
\WF@size=\dimen176
\c@WF@wrappedlines=\count287
\WF@box=\box68
```

```
\WF@everypar=\toks24
Package: wrapfig 2003/01/31  v 3.6
)
\wraplines=\count288
\@authorfigbox=\box69
\@authorfigboxdim=\skip57
\biofigadjskip=\skip58
(c:/TeXLive/2022/texmf-dist/tex/latex/natbib/natbib.sty
Package: natbib 2010/09/13 8.31b (PWD, AO)
\bibhang=\skip59
\bibsep=\skip60
LaTeX Info: Redefining \cite on input line 694.
\c@NAT@ctr=\count289
)
Package hyperref Info: Option `colorlinks' set `false' on input line
1725.
) (c:/TeXLive/2022/texmf-dist/tex/latex/multirow/multirow.sty
Package: multirow 2021/03/15 v2.8 Span multiple rows of a table
\multirow@colwidth=\skip61
\multirow@cntb=\count290
\multirow@dima=\skip62
\bigstrutjot=\dimen177
) (c:/TeXLive/2022/texmf-dist/tex/latex/amsmath/amsmath.sty
Package: amsmath 2022/04/08 v2.17n AMS math features
\@mathmargin=\skip63
For additional information on amsmath, use the `?' option.
(c:/TeXLive/2022/texmf-dist/tex/latex/amsmath/amstext.sty
Package: amstext 2021/08/26 v2.01 AMS text
(c:/TeXLive/2022/texmf-dist/tex/latex/amsmath/amsgen.sty
File: amsgen.sty 1999/11/30 v2.0 generic functions
\@emptytoks=\toks25
\ex@=\dimen178
)) (c:/TeXLive/2022/texmf-dist/tex/latex/amsmath/amsbsy.sty
Package: amsbsy 1999/11/29 v1.2d Bold Symbols
\pmbraise@=\dimen179
) (c:/TeXLive/2022/texmf-dist/tex/latex/amsmath/amsopn.sty
Package: amsopn 2022/04/08 v2.04 operator names
)
\inf@bad=\count291
LaTeX Info: Redefining \frac on input line 234.
\uproot@=\count292
\leftroot@=\count293
LaTeX Info: Redefining \overline on input line 399.
LaTeX Info: Redefining \colon on input line 410.
\classnum@=\count294
\DOTSCASE@=\count295
LaTeX Info: Redefining \ldots on input line 496.
LaTeX Info: Redefining \dots on input line 499.
LaTeX Info: Redefining \cdots on input line 620.
\Mathstrutbox@=\box70
\strutbox@=\box71
LaTeX Info: Redefining \big on input line 722.
LaTeX Info: Redefining \Big on input line 723.
LaTeX Info: Redefining \bigg on input line 724.
```

```
LaTeX Info: Redefining \Bigg on input line 725.
\big@size=\dimen180
LaTeX Font Info:    Redeclaring font encoding OML on input line 743.
LaTeX Font Info:    Redeclaring font encoding OMS on input line 744.
\macc@depth=\count296
LaTeX Info: Redefining \bmod on input line 905.
LaTeX Info: Redefining \pmod on input line 910.
LaTeX Info: Redefining \smash on input line 940.
LaTeX Info: Redefining \relbar on input line 970.
LaTeX Info: Redefining \Relbar on input line 971.
\c@MaxMatrixCols=\count297
\dotsspace@=\muskip17
\c@parentequation=\count298
\dspbrk@lvl=\count299
\tag@help=\toks26
\row@=\count300
\column@=\count301
\maxfields@=\count302
\andhelp@=\toks27
\eqnshift@=\dimen181
\alignsep@=\dimen182
\tagshift@=\dimen183
\tagwidth@=\dimen184
\totwidth@=\dimen185
\lineht@=\dimen186
\@envbody=\toks28
\multlinegap=\skip64
\multlinetaggap=\skip65
\mathdisplay@stack=\toks29
LaTeX Info: Redefining \[ on input line 2953.
LaTeX Info: Redefining \] on input line 2954.
) (c:/TeXLive/2022/texmf-dist/tex/latex/amsfonts/amssymb.sty
Package: amssymb 2013/01/14 v3.01 AMS font symbols
(c:/TeXLive/2022/texmf-dist/tex/latex/amsfonts/amsfonts.sty
Package: amsfonts 2013/01/14 v3.01 Basic AMSFonts support
\symAMSa=\mathgroup4
\symAMSb=\mathgroup5
LaTeX Font Info:    Redeclaring math symbol \hbar on input line 98.
LaTeX Font Info:    Overwriting math alphabet `\mathfrak' in version
`bold'
(Font)                  U/euf/m/n --> U/euf/b/n on input line 106.
)) (c:/TeXLive/2022/texmf-dist/tex/latex/amscls/amsthm.sty
Package: amsthm 2020/05/29 v2.20.6
\thm@style=\toks30
\thm@bodyfont=\toks31
\thm@headfont=\toks32
\thm@notefont=\toks33
\thm@headpunct=\toks34
\thm@preskip=\skip66
\thm@postskip=\skip67
\thm@headsep=\skip68
\dth@everypar=\toks35
) (c:/TeXLive/2022/texmf-dist/tex/latex/jknapltx/mathrsfs.sty
Package: mathrsfs 1996/01/01 Math RSFS package v1.0 (jk)
```

```
\symrsfs=\mathgroup6
) (c:/TeXLive/2022/texmf-dist/tex/latex/xcolor/xcolor.sty
Package: xcolor 2022/06/12 v2.14 LaTeX color extensions (UK)
(c:/TeXLive/2022/texmf-dist/tex/latex/graphics-cfg/color.cfg
File: color.cfg 2016/01/02 v1.6 sample color configuration
)
Package xcolor Info: Driver file: pdftex.def on input line 227.
(c:/TeXLive/2022/texmf-dist/tex/latex/graphics/mathcolor.ltx)
Package xcolor Info: Model `cmy' substituted by `cmy0' on input line
1353.
Package xcolor Info: Model `hsb' substituted by `rgb' on input line 1357.
Package xcolor Info: Model `RGB' extended on input line 1369.
Package xcolor Info: Model `HTML' substituted by `rgb' on input line
1371.
Package xcolor Info: Model `Hsb' substituted by `hsb' on input line 1372.
Package xcolor Info: Model `tHsb' substituted by `hsb' on input line
1373.
Package xcolor Info: Model `HSB' substituted by `hsb' on input line 1374.
Package xcolor Info: Model `Gray' substituted by `gray' on input line
1375.
Package xcolor Info: Model `wave' substituted by `hsb' on input line
1376.
) (c:/TeXLive/2022/texmf-dist/tex/latex/base/textcomp.sty
Package: textcomp 2020/02/02 v2.0n Standard LaTeX package
) (c:/TeXLive/2022/texmf-dist/tex/latex/ncctools/manyfoot.sty
Package: manyfoot 2019/08/03 v1.11 Many Footnote Levels Package (NCC)
(c:/TeXLive/2022/texmf-dist/tex/latex/ncctools/nccfoots.sty
Package: nccfoots 2005/02/03 v1.2 NCC Footnotes Package (NCC)
)
\MFL@columnwidth=\dimen187
) (c:/TeXLive/2022/texmf-dist/tex/latex/booktabs/booktabs.sty
Package: booktabs 2020/01/12 v1.61803398 Publication quality tables
\heavyrulewidth=\dimen188
\lightrulewidth=\dimen189
\cmidrulewidth=\dimen190
\belowrulesep=\dimen191
\belowbottomsep=\dimen192
\aboverulesep=\dimen193
\abovetopsep=\dimen194
\cmidrulesep=\dimen195
\cmidrulekern=\dimen196
\defaultaddspace=\dimen197
\@cmidla=\count303
\@cmidlb=\count304
\@aboverulesep=\dimen198
\@belowrulesep=\dimen199
\@thisruleclass=\count305
\@lastruleclass=\count306
\@thisrulewidth=\dimen256
) (c:/TeXLive/2022/texmf-dist/tex/latex/algorithms/algorithm.sty
Package: algorithm 2009/08/24 v0.1 Document Style `algorithm' - floating
enviro
nment
(c:/TeXLive/2022/texmf-dist/tex/latex/float/float.sty
```

```
Package: float 2001/11/08 v1.3d Float enhancements (AL)
\c@float@type=\count307
\float@exts=\toks36
\float@box=\box72
\@float@everytoks=\toks37
\@floatcapt=\box73
)
\@float@every@algorithm=\toks38
\c@algorithm=\count308
) (c:/TeXLive/2022/texmf-dist/tex/latex/algorithmicx/algorithmicx.sty
Package: algorithmicx 2005/04/27 v1.2 Algorithmicx
Document Style algorithmicx 1.2 - a greatly improved `algorithmic' style
\c@ALG@line=\count309
\c@ALG@rem=\count310
\c@ALG@nested=\count311
\ALG@tlm=\skip69
\ALG@thistlm=\skip70
\c@ALG@Lnr=\count312
\c@ALG@blocknr=\count313
\c@ALG@storecount=\count314
\c@ALG@tmpcounter=\count315
\ALG@tmplength=\skip71
) (c:/TeXLive/2022/texmf-dist/tex/latex/algorithmicx/algpseudocode.sty
Package: algpseudocode
Document Style - pseudocode environments for use with the `algorithmicx'
style
) (c:/TeXLive/2022/texmf-dist/tex/latex/listings/listings.sty
\lst@mode=\count316
\lst@gtempboxa=\box74
\lst@token=\toks39
\lst@length=\count317
\lst@currlwidth=\dimen257
\lst@column=\count318
\lst@pos=\count319
\lst@lostspace=\dimen258
\lst@width=\dimen259
\lst@newlines=\count320
\lst@lineno=\count321
\lst@maxwidth=\dimen260
(c:/TeXLive/2022/texmf-dist/tex/latex/listings/lstmisc.sty
File: lstmisc.sty 2023/02/27 1.9 (Carsten Heinz)
\c@lstnumber=\count322
\lst@skipnumbers=\count323
\lst@framebox=\box75
) (c:/TeXLive/2022/texmf-dist/tex/latex/listings/listings.cfg
File: listings.cfg 2023/02/27 1.9 listings configuration
))
Package: listings 2023/02/27 1.9 (Carsten Heinz)

Package amsthm Warning: Unknown theoremstyle `thmstyleone' on input line
77.

\c@theorem=\count324
```

```
Package amsthm Warning: Unknown theoremstyle `thmstyletwo' on input line
84.

\c@example=\count325
\c@remark=\count326

Package amsthm Warning: Unknown theoremstyle `thmstylethree' on input
line 88.

\c@definition=\count327
(c:/TeXLive/2022/texmf-dist/tex/latex/l3backend/l3backend-pdftex.def
File: l3backend-pdftex.def 2023-01-16 L3 backend support: PDF output
(pdfTeX)
\l__color_backend_stack_int=\count328
\l__pdf_internal_box=\box76
) (./sn-article.aux)
\openout1 = `sn-article.aux'.

LaTeX Font Info:    Checking defaults for OML/cmm/m/it on input line 94.
LaTeX Font Info:    ... okay on input line 94.
LaTeX Font Info:    Checking defaults for OMS/cmsy/m/n on input line 94.
LaTeX Font Info:    ... okay on input line 94.
LaTeX Font Info:    Checking defaults for OT1/cmr/m/n on input line 94.
LaTeX Font Info:    ... okay on input line 94.
LaTeX Font Info:    Checking defaults for T1/cmr/m/n on input line 94.
LaTeX Font Info:    ... okay on input line 94.
LaTeX Font Info:    Checking defaults for TS1/cmr/m/n on input line 94.
LaTeX Font Info:    ... okay on input line 94.
LaTeX Font Info:    Checking defaults for OMX/cmex/m/n on input line 94.
LaTeX Font Info:    ... okay on input line 94.
LaTeX Font Info:    Checking defaults for U/cmr/m/n on input line 94.
LaTeX Font Info:    ... okay on input line 94.
LaTeX Font Info:    Checking defaults for PD1/pdf/m/n on input line 94.
LaTeX Font Info:    ... okay on input line 94.
LaTeX Font Info:    Checking defaults for PU/pdf/m/n on input line 94.
LaTeX Font Info:    ... okay on input line 94.
\footinsA=\insert252
\c@footnoteA=\count329
*geometry* driver: auto-detecting
*geometry* detected driver: pdftex
*geometry* verbose mode - [ preamble ] result:
* driver: pdftex
* paper: custom
* layout: <same size as paper>
* layoutoffset:(h,v)=(0.0pt,0.0pt)
* bindingoffset: 17.07164pt
* modes: twoside
* h-part:(L,W,R)=(83.37448pt, 372.0pt, 125.06175pt)
* v-part:(T,H,B)=(73.97716pt, 552.69478pt, 218.37491pt)
* \paperwidth=597.50787pt
* \paperheight=845.04684pt
* \textwidth=372.0pt
* \textheight=552.69478pt
* \oddsidemargin=28.17613pt
```

```
* \evensidemargin=52.79176pt
* \topmargin=-19.72638pt
* \headheight=5.5pt
* \headsep=15.93355pt
* \topskip=10.0pt
* \footskip=28.45274pt
* \marginparwidth=34.1433pt
* \marginparsep=14.22636pt
* \columnsep=14.22636pt
* \skip\footins=8.0pt plus 4.0pt
* \hoffset=0.0pt
* \voffset=0.0pt
* \mag=1000
* \@twocolumnfalse
* \@twosidetrue
* \@mparswitchtrue
* \@reversemargintrue
* (1in=72.27pt=25.4mm, 1cm=28.453pt)

(c:/TeXLive/2022/texmf-dist/tex/context/base/mkii/supp-pdf.mkii
[Loading MPS to PDF converter (version 2006.09.02).]
\scratchcounter=\count330
\scratchdimen=\dimen261
\scratchbox=\box77
\nofMPsegments=\count331
\nofMParguments=\count332
\everyMPshowfont=\toks40
\MPscratchCnt=\count333
\MPscratchDim=\dimen262
\MPnumerator=\count334
\makeMPintoPDFobject=\count335
\everyMPtoPDFconversion=\toks41
) (c:/TeXLive/2022/texmf-dist/tex/latex/epstopdf-pkg/epstopdf-base.sty
Package: epstopdf-base 2020-01-24 v2.11 Base part for package epstopdf
Package epstopdf-base Info: Redefining graphics rule for `.eps' on input
line 4
85.
(c:/TeXLive/2022/texmf-dist/tex/latex/latexconfig/epstopdf-sys.cfg
File: epstopdf-sys.cfg 2010/07/13 v1.3 Configuration of (r)epstopdf for
TeX Liv
e
))
Package hyperref Info: Link coloring OFF on input line 94.
(./sn-article.out) (./sn-article.out)
\@outlinefile=\write3
\openout3 = `sn-article.out'.

\c@lstlisting=\count336
LaTeX Font Info:    Calculating math sizes for size <12.045> on input
line 146.


LaTeX Font Warning: Font shape `OT1/cmr/bx/n' in size <8.43146> not
available
```

```
(Font)              size <8> substituted on input line 146.


LaTeX Font Warning: Font shape `OML/cmm/b/it' in size <8.43146> not
available
(Font)              size <8> substituted on input line 146.


LaTeX Font Warning: Font shape `OMS/cmsy/b/n' in size <8.43146> not
available
(Font)              size <8> substituted on input line 146.

LaTeX Font Info:    Trying to load font information for U+msa on input
line 146
.
(c:/TeXLive/2022/texmf-dist/tex/latex/amsfonts/umsa.fd
File: umsa.fd 2013/01/14 v3.01 AMS symbols A
)
LaTeX Font Info:    Trying to load font information for U+msb on input
line 146
.
(c:/TeXLive/2022/texmf-dist/tex/latex/amsfonts/umsb.fd
File: umsb.fd 2013/01/14 v3.01 AMS symbols B
)
LaTeX Font Info:    Trying to load font information for U+rsfs on input
line 14
6.
(c:/TeXLive/2022/texmf-dist/tex/latex/jknapltx/ursfs.fd
File: ursfs.fd 1998/03/24 rsfs font definition file (jk)
)

LaTeX Font Warning: Font shape `U/rsfs/m/n' in size <8.43146> not
available
(Font)              size <8> substituted on input line 146.


LaTeX Font Warning: Font shape `OT1/cmr/m/n' in size <8.43146> not
available
(Font)              size <8> substituted on input line 146.

LaTeX Font Info:    Calculating math sizes for size <11.04124> on input
line 14
6.

LaTeX Font Warning: Font shape `OT1/cmr/bx/n' in size <5.52061> not
available
(Font)              size <6> substituted on input line 146.


LaTeX Font Warning: Font shape `OML/cmm/b/it' in size <5.52061> not
available
(Font)              size <6> substituted on input line 146.
```

```
LaTeX Font Warning: Font shape `OMS/cmsy/b/n' in size <5.52061> not
available
(Font)              size <6> substituted on input line 146.


LaTeX Font Warning: Font shape `U/rsfs/m/n' in size <5.52061> not
available
(Font)              size <6> substituted on input line 146.

LaTeX Font Info:    Calculating math sizes for size <10.03749> on input
line 14
6.

Package natbib Warning: Citation `breiman_statistical_2001' on page 1
undefined
 on input line 153.


Package natbib Warning: Citation `shmueli_explain_2010' on page 1
undefined on
input line 154.


Package natbib Warning: Citation `breiman_random_2001' on page 1
undefined on i
nput line 167.


Package natbib Warning: Citation `boser_training_1992' on page 1
undefined on i
nput line 167.


Package natbib Warning: Citation `anderson_introduction_1995' on page 1
undefin
ed on input line 167.


Underfull \vbox (badness 10000) has occurred while \output is active []

[1

{c:/TeXLive/2022/texmf-var/fonts/map/pdftex/updmap/pdftex.map}]

Package natbib Warning: Citation `stahl-ethics' on page 2 undefined on
input li
ne 177.


Package natbib Warning: Citation `adadi_peeking_2018' on page 2 undefined
on in
put line 186.
```

Package natbib Warning: Citation `arrieta_explainable_2020' on page 2
undefined
 on input line 186.

Package natbib Warning: Citation `asimov_grand_1985' on page 2 undefined
on inp
ut line 197.

Package natbib Warning: Citation `buja_grand_1986' on page 2 undefined on
input
 line 197.

Package natbib Warning: Citation `lee_state_2021' on page 2 undefined on
input
line 197.

Package natbib Warning: Citation `wickham_visualizing_2015' on page 2
undefined
 on input line 200.

Package natbib Warning: Citation `cook_interactive_2007' on page 2
undefined on
 input line 201.

Package natbib Warning: Citation `Caragea2008' on page 2 undefined on
input lin
e 204.

Package natbib Warning: Citation `lee_pptree_2013' on page 2 undefined on
input
 line 204.

Package natbib Warning: Citation `da_silva_projection_2021' on page 2
undefined
 on input line 204.

Package natbib Warning: Citation `cook_manual_1997' on page 2 undefined
on inpu
t line 206.

Package natbib Warning: Citation `spyrison_spinifex_2020' on page 2
undefined o
n input line 206.

```
Package natbib Warning: Citation `biecek_ceterisparibus_2020' on page 2
undefin
ed on input line 211.


Underfull \vbox (badness 10000) has occurred while \output is active []

[2]

Package natbib Warning: Citation `arrieta_explainable_2020' on page 3
undefined
 on input line 234.


Package natbib Warning: Citation `shrikumar_not_2016' on page 3 undefined
on in
put line 236.


Package natbib Warning: Citation `shrikumar_learning_2017' on page 3
undefined
on input line 236.


Package natbib Warning: Citation `ribeiro_why_2016' on page 3 undefined
on inpu
t line 238.


Package natbib Warning: Citation `lundberg_unified_2017' on page 3
undefined on
 input line 239.


Package natbib Warning: Citation `simonyan_deep_2014' on page 3 undefined
on in
put line 245.


Package natbib Warning: Citation `vanni_textual_2018' on page 3 undefined
on in
put line 249.


Package natbib Warning: Citation `ribeiro_why_2016' on page 3 undefined
on inpu
t line 252.


Package natbib Warning: Citation `shapley_value_1953' on page 3 undefined
on in
put line 258.
```

```
Package natbib Warning: Citation `strumbelj_efficient_2010' on page 3
undefined
 on input line 261.


Package natbib Warning: Citation `molnar2022' on page 3 undefined on
input line
 274.


! LaTeX Error: File `figures/shap_distr_bd' not found.

See the LaTeX manual or LaTeX Companion for explanation.
Type  H <return>  for immediate help.
 ...

l.281 ...th=0.85\linewidth]{figures/shap_distr_bd}

I could not locate the file with any of these extensions:
.pdf,.png,.jpg,.mps,.jpeg,.jbig2,.jb2,.PDF,.PNG,.JPG,.JPEG,.JBIG2,.JB2,.e
ps
Try typing  <return>  to proceed.
If that doesn't work, type  X <return>  to quit.


LaTeX Warning: `!h' float specifier changed to `!ht'.


Package natbib Warning: Citation `leone_fifa_2020' on page 3 undefined on
input
 line 288.


Package natbib Warning: Citation `biecek_explanatory_2021' on page 3
undefined
on input line 289.


Package natbib Warning: Citation `gosiewska_ibreakdown_2019' on page 3
undefine
d on input line 301.


Underfull \vbox (badness 10000) has occurred while \output is active []

[3]

Package natbib Warning: Citation `wickham_visualizing_2015' on page 4
undefined
 on input line 331.
```

Package natbib Warning: Citation `lundberg_consistent_2018' on page 4
undefined
 on input line 342.

Package natbib Warning: Citation `kominsarczyk_treeshap_2023' on page 4
undefin
ed on input line 349.

Package natbib Warning: Citation `fastshap' on page 4 undefined on input
line 3
49.

Package natbib Warning: Citation `kernelshap' on page 4 undefined on
input line
 350.

Package natbib Warning: Citation `shapr' on page 4 undefined on input
line 350.

Package natbib Warning: Citation `shapviz' on page 4 undefined on input
line 35
1.

Package natbib Warning: Citation `PPtreeregViz' on page 4 undefined on
input li
ne 352.

Package natbib Warning: Citation `ExplainPrediction' on page 4 undefined
on inp
ut line 353.

Package natbib Warning: Citation `flashlight' on page 4 undefined on
input line
 353.

Package natbib Warning: Citation `biecek_dalex_2018' on page 4 undefined
on inp
ut line 354.

Package natbib Warning: Citation `molnar2022' on page 4 undefined on
input line
 355.

```
Underfull \vbox (badness 10000) has occurred while \output is active []

[4
pdfTeX warning (ext4): destination with the same identifier
(name{figure.1}) ha
s been already used, duplicate ignored
<argument> ...shipout:D \box_use:N \l_shipout_box

\__shipout_drop_firstpage_...
l.357
      ]

Package natbib Warning: Citation `lee_state_2021' on page 5 undefined on
input
line 368.


Package natbib Warning: Citation `cook_grand_2008' on page 5 undefined on
input
 line 368.


Package natbib Warning: Citation `cook_manual_1997' on page 5 undefined
on inpu
t line 370.


Package natbib Warning: Citation `karwowski_international_2006' on page 5
undef
ined on input line 379.


Package natbib Warning: Citation `spyrison_spinifex_2020' on page 5
undefined o
n input line 382.


! LaTeX Error: File `./figures/radial_tour' not found.

See the LaTeX manual or LaTeX Companion for explanation.
Type  H <return>  for immediate help.
 ...

l.390 ...th=0.99\linewidth]{./figures/radial_tour}

I could not locate the file with any of these extensions:
.pdf,.png,.jpg,.mps,.jpeg,.jbig2,.jb2,.PDF,.PNG,.JPG,.JPEG,.JBIG2,.JB2,.e
ps
Try typing  <return>  to proceed.
If that doesn't work, type  X <return>  to quit.
```

```
Package natbib Warning: Citation `roberts_state_2007' on page 5 undefined
on in
put line 398.


Package natbib Warning: Citation `unwin_ensemble_2018' on page 5
undefined on i
nput line 399.


Underfull \vbox (badness 10000) has occurred while \output is active []

[5
pdfTeX warning (ext4): destination with the same identifier
(name{figure.2}) ha
s been already used, duplicate ignored
<argument> ...shipout:D \box_use:N \l_shipout_box

\__shipout_drop_firstpage_...
l.412 \subsection{Global View}\label{global-view}}
                                                  ]
Underfull \vbox (badness 10000) has occurred while \output is active []

[6]

Package natbib Warning: Citation `ocagne_coordonnees_1885' on page 7
undefined
on input line 461.


! LaTeX Error: File `./figures/app_classification' not found.

See the LaTeX manual or LaTeX Companion for explanation.
Type  H <return>  for immediate help.
 ...

l.512 ...\linewidth]{./figures/app_classification}

I could not locate the file with any of these extensions:
.pdf,.png,.jpg,.mps,.jpeg,.jbig2,.jb2,.PDF,.PNG,.JPG,.JPEG,.JBIG2,.JB2,.e
ps
Try typing  <return>  to proceed.
If that doesn't work, type  X <return>  to quit.


Underfull \vbox (badness 10000) has occurred while \output is active []

[7]

! LaTeX Error: File `./figures/app_regression_interactions' not found.

See the LaTeX manual or LaTeX Companion for explanation.
Type  H <return>  for immediate help.
 ...
```

```
l.547 ...h]{./figures/app_regression_interactions}

I could not locate the file with any of these extensions:
.pdf,.png,.jpg,.mps,.jpeg,.jbig2,.jb2,.PDF,.PNG,.JPG,.JPEG,.JBIG2,.JB2,.e
ps
Try typing  <return>  to proceed.
If that doesn't work, type  X <return>  to quit.


Underfull \vbox (badness 10000) has occurred while \output is active []

[8
pdfTeX warning (ext4): destination with the same identifier
(name{figure.3}) ha
s been already used, duplicate ignored
<argument> ...shipout:D \box_use:N \l_shipout_box

\__shipout_drop_firstpage_...
l.563

pdfTeX warning (ext4): destination with the same identifier
(name{figure.4}) ha
s been already used, duplicate ignored
<argument> ...shipout:D \box_use:N \l_shipout_box

\__shipout_drop_firstpage_...
l.563
      ]
Underfull \vbox (badness 10000) has occurred while \output is active []

[9]
! Undefined control sequence.
<recently read> \tightlist

l.585 \tightlist

The control sequence at the end of the top line
of your error message was never \def'ed. If you have
misspelled it (e.g., `\hobx'), type `I' and the correct
spelling (e.g., `I\hbox'). Otherwise just continue,
and I'll forget about whatever was undefined.


Package natbib Warning: Citation `liaw_classification_2002' on page 10
undefine
d on input line 593.


Package natbib Warning: Citation `kominsarczyk_treeshap_2023' on page 10
undefi
ned on input line 602.
```

```
Package natbib Warning: Citation `greenwell_fastshap_2020' on page 10
undefined
 on input line 624.


Package natbib Warning: Citation `gorman_ecological_2014' on page 10
undefined
on input line 637.


Package natbib Warning: Citation `horst_palmerpenguins_2020' on page 10
undefin
ed on input line 637.


! LaTeX Error: File `./figures/case_penguins' not found.

See the LaTeX manual or LaTeX Companion for explanation.
Type  H <return>  for immediate help.
 ...

l.648 ...dth=1\linewidth]{./figures/case_penguins}

I could not locate the file with any of these extensions:
.pdf,.png,.jpg,.mps,.jpeg,.jbig2,.jb2,.PDF,.PNG,.JPG,.JPEG,.JBIG2,.JB2,.e
ps
Try typing  <return>  to proceed.
If that doesn't work, type  X <return>  to quit.


Underfull \vbox (badness 10000) has occurred while \output is active []

[10

]

! LaTeX Error: File `./figures/case_penguins_BlFl' not found.

See the LaTeX manual or LaTeX Companion for explanation.
Type  H <return>  for immediate help.
 ...

l.664 ...\linewidth]{./figures/case_penguins_BlFl}

I could not locate the file with any of these extensions:
.pdf,.png,.jpg,.mps,.jpeg,.jbig2,.jb2,.PDF,.PNG,.JPG,.JPEG,.JBIG2,.JB2,.e
ps
Try typing  <return>  to proceed.
If that doesn't work, type  X <return>  to quit.


Underfull \vbox (badness 10000) has occurred while \output is active []

[11
```

```
pdfTeX warning (ext4): destination with the same identifier
(name{figure.5}) ha
s been already used, duplicate ignored
<argument> ...shipout:D \box_use:N \l_shipout_box

\__shipout_drop_firstpage_...
l.709 ...abel{chocolates-milkdark-classification}}

pdfTeX warning (ext4): destination with the same identifier
(name{figure.6}) ha
s been already used, duplicate ignored
<argument> ...shipout:D \box_use:N \l_shipout_box

\__shipout_drop_firstpage_...
l.709 ...abel{chocolates-milkdark-classification}}
                                                  ]

! LaTeX Error: File `./figures/case_chocolates' not found.

See the LaTeX manual or LaTeX Companion for explanation.
Type  H <return>  for immediate help.
 ...

l.728 ...h=1\linewidth]{./figures/case_chocolates}

I could not locate the file with any of these extensions:
.pdf,.png,.jpg,.mps,.jpeg,.jbig2,.jb2,.PDF,.PNG,.JPG,.JPEG,.JBIG2,.JB2,.e
ps
Try typing  <return>  to proceed.
If that doesn't work, type  X <return>  to quit.


Underfull \vbox (badness 10000) has occurred while \output is active []

[12
pdfTeX warning (ext4): destination with the same identifier
(name{figure.7}) ha
s been already used, duplicate ignored
<argument> ...shipout:D \box_use:N \l_shipout_box

\__shipout_drop_firstpage_...
l.767 F
       rom the density plot, which is the attribution projection
corresponding]

! LaTeX Error: File `./figures/case_chocolates_inverse' not found.

See the LaTeX manual or LaTeX Companion for explanation.
Type  H <return>  for immediate help.
 ...

l.776 ...width]{./figures/case_chocolates_inverse}

I could not locate the file with any of these extensions:
```

```
.pdf,.png,.jpg,.mps,.jpeg,.jbig2,.jb2,.PDF,.PNG,.JPG,.JPEG,.JBIG2,.JB2,.e
ps
Try typing  <return>  to proceed.
If that doesn't work, type  X <return>  to quit.


Package natbib Warning: Citation `leone_fifa_2020' on page 13 undefined
on inpu
t line 790.


Package natbib Warning: Citation `biecek_dalex_2018' on page 13 undefined
on in
put line 790.


! LaTeX Error: File `./figures/case_fifa' not found.

See the LaTeX manual or LaTeX Companion for explanation.
Type  H <return>  for immediate help.
 ...

l.802 ...width=0.9\linewidth]{./figures/case_fifa}

I could not locate the file with any of these extensions:
.pdf,.png,.jpg,.mps,.jpeg,.jbig2,.jb2,.PDF,.PNG,.JPG,.JPEG,.JBIG2,.JB2,.e
ps
Try typing  <return>  to proceed.
If that doesn't work, type  X <return>  to quit.


LaTeX Warning: `!h' float specifier changed to `!ht'.


Underfull \vbox (badness 10000) has occurred while \output is active []

[13
pdfTeX warning (ext4): destination with the same identifier
(name{figure.8}) ha
s been already used, duplicate ignored
<argument> ...shipout:D \box_use:N \l_shipout_box

\__shipout_drop_firstpage_...
l.814 ... variable with low importance is varied.}
                                                  ]

! LaTeX Error: File `./figures/case_ames2018' not found.

See the LaTeX manual or LaTeX Companion for explanation.
Type  H <return>  for immediate help.
 ...

l.818 ...h=0.9\linewidth]{./figures/case_ames2018}
```

```
I could not locate the file with any of these extensions:
.pdf,.png,.jpg,.mps,.jpeg,.jbig2,.jb2,.PDF,.PNG,.JPG,.JPEG,.JBIG2,.JB2,.e
ps
Try typing  <return>  to proceed.
If that doesn't work, type  X <return>  to quit.


Package natbib Warning: Citation `de_cock_ames_2011' on page 14 undefined
on in
put line 829.


Underfull \vbox (badness 10000) has occurred while \output is active []

[14
pdfTeX warning (ext4): destination with the same identifier
(name{figure.9}) ha
s been already used, duplicate ignored
<argument> ...shipout:D \box_use:N \l_shipout_box

\__shipout_drop_firstpage_...
l.844 ...n{Discussion}\label{sec:cheemdiscussion}}

pdfTeX warning (ext4): destination with the same identifier
(name{figure.10}) h
as been already used, duplicate ignored
<argument> ...shipout:D \box_use:N \l_shipout_box

\__shipout_drop_firstpage_...
l.844 ...n{Discussion}\label{sec:cheemdiscussion}}
                                                  ]
Underfull \vbox (badness 10000) has occurred while \output is active []

[15]

Package natbib Warning: Citation `spyrison_cheem_2023' on page 16
undefined on
input line 893.


Package natbib Warning: Citation `greenwell_gbm_2020' on page 16
undefined on i
nput line 898.


Package natbib Warning: Citation `shi_lightgbm_2022' on page 16 undefined
on in
put line 899.


Package natbib Warning: Citation `liaw_classification_2002' on page 16
undefine
d on input line 899.
```

```
Package natbib Warning: Citation `wright_ranger_2017' on page 16
undefined on i
nput line 900.


Package natbib Warning: Citation `chen_xgboost_2021' on page 16 undefined
on in
put line 900.


Package natbib Warning: Citation `biecek_dalex_2018' on page 16 undefined
on in
put line 903.


Package natbib Warning: Citation `biecek_explanatory_2021' on page 16
undefined
 on input line 903.


Package natbib Warning: Citation `chang_shiny_2021' on page 16 undefined
on inp
ut line 905.


Package natbib Warning: Citation `spyrison_spinifex_2020' on page 16
undefined
on input line 907.


Package natbib Warning: Citation `wickham_ggplot2_2016' on page 16
undefined on
 input line 908.


Package natbib Warning: Citation `sievert_interactive_2020' on page 16
undefine
d on input line 910.


Package natbib Warning: Citation `biecek_dalex_2018' on page 16 undefined
on in
put line 911.


Package natbib Warning: Citation `biecek_explanatory_2021' on page 16
undefined
 on input line 912.


Package hyperref Warning: Difference (4) between bookmark levels is
greater
(hyperref)                than one, level fixed on input line 934.
```

```
No file sn-article.bbl.

Package natbib Warning: There were undefined citations.


Underfull \vbox (badness 10000) has occurred while \output is active []

[16] (./sn-article.aux)

LaTeX Font Warning: Size substitutions with differences
(Font)              up to 0.47939pt have occurred.


LaTeX Warning: Label(s) may have changed. Rerun to get cross-references
right.

Package rerunfilecheck Info: File `sn-article.out' has not changed.
(rerunfilecheck)             Checksum:
5C5BCD0AC6371660644804559E4158CB;2750.
 )
Here is how much of TeX's memory you used:
 14756 strings out of 476024
 232632 string characters out of 5794017
 1862382 words of memory out of 5000000
 34864 multiletter control sequences out of 15000+600000
 529895 words of font info for 104 fonts, out of 8000000 for 9000
 1151 hyphenation exceptions out of 8191
 90i,8n,90p,1164b,473s stack positions out of
10000i,1000n,20000p,200000b,200000s
{c:/TeXLive/2022/texmf-dist/fonts/enc/dvips/cm-super/cm-super-
ts1.enc}<c:/TeX
Live/2022/texmf-
dist/fonts/type1/public/amsfonts/cm/cmbx10.pfb><c:/TeXLive/2022
/texmf-
dist/fonts/type1/public/amsfonts/cm/cmbx12.pfb><c:/TeXLive/2022/texmf-di
st/fonts/type1/public/amsfonts/cm/cmbx8.pfb><c:/TeXLive/2022/texmf-
dist/fonts/t
ype1/public/amsfonts/cm/cmbx9.pfb><c:/TeXLive/2022/texmf-
dist/fonts/type1/publi
c/amsfonts/cm/cmmi10.pfb><c:/TeXLive/2022/texmf-
dist/fonts/type1/public/amsfont
s/cm/cmmi8.pfb><c:/TeXLive/2022/texmf-
dist/fonts/type1/public/amsfonts/cm/cmr10
.pfb><c:/TeXLive/2022/texmf-
dist/fonts/type1/public/amsfonts/cm/cmr12.pfb><c:/T
eXLive/2022/texmf-
dist/fonts/type1/public/amsfonts/cm/cmr17.pfb><c:/TeXLive/202
2/texmf-
dist/fonts/type1/public/amsfonts/cm/cmr8.pfb><c:/TeXLive/2022/texmf-dis
t/fonts/type1/public/amsfonts/cm/cmr9.pfb><c:/TeXLive/2022/texmf-
dist/fonts/typ
e1/public/amsfonts/cm/cmss8.pfb><c:/TeXLive/2022/texmf-
dist/fonts/type1/public/
```

```
amsfonts/cm/cmsy10.pfb><c:/TeXLive/2022/texmf-
dist/fonts/type1/public/amsfonts/
cm/cmti10.pfb><c:/TeXLive/2022/texmf-
dist/fonts/type1/public/amsfonts/cm/cmtt10
.pfb><c:/TeXLive/2022/texmf-dist/fonts/type1/public/cm-
super/sfrm1000.pfb>
Output written on sn-article.pdf (16 pages, 270086 bytes).
PDF statistics:
 311 PDF objects out of 1000 (max. 8388607)
 258 compressed objects within 3 object streams
 63 named destinations out of 1000 (max. 500000)
 145 words of extra memory for PDF output out of 10000 (max. 10000000)
```

Exploring Local Explanations of Nonlinear Models Using Animated Linear Projections

Nicholas Spyrison[1*], Dianne Cook[2] and Przemyslaw Biecek[3]

[1*]Faculty of Information Technology, Monash University, Wellington Rd, Clayton, VIC 3800, Australia.
[2]Dept of Econometrics and Business Statistics, Monash University, Wellington Rd, Clayton, VIC 3800, Australia.
[3]Faculty of Mathematics and Information Science, Warsaw University of Technology, Koszykowa 75, Warsaw, 00-662, Poland.

*Corresponding author(s). E-mail(s): spyrison@gmail.com;
Contributing authors: dicook@monash.edu;
przemyslaw.biecek@pw.edu.pl;

Abstract

The increased predictive power of machine learning models comes at the cost of increased complexity and loss of interpretability, particularly in comparison to parametric statistical models. This trade-off has led to the emergence of eXplainable AI (XAI) which provides methods, such as local explanations (LEs) and local variable attributions (LVAs), to shed light on how a model use predictors to arrive at a prediction. These provide a point estimate of the linear variable importance in the vicinity of a single observation. However, LVAs tend not to effectively handle association between predictors. To understand how the interaction between predictors affects the variable importance estimate, we can convert LVAs into linear projections and use the radial tour. This is also useful for learning how a model has made a mistake, or the effect of outliers, or the clustering of observations. The approach is illustrated with examples from categorical (penguin species, chocolate types) and quantitative (soccer/football salaries, house prices) response models. The methods are implemented in the R package cheem, available on CRAN.

Keywords: explainable artificial intelligence, nonlinear model interpretability, visual analytics, local explanations, grand tour, radial tour

1 Introduction

There are different reasons and purposes for fitting a model. According to the taxonomies of Breiman (2001b) and Shmueli (2010), it can be useful to group models into two types: explanatory and predictive. Explanatory modeling is used for inferential purposes, while predictive modeling focuses solely on the performance of an objective function. The intended use of the model has important implications for its selection and development. Interpretability is critical in explanatory modeling to draw meaningful inferential conclusions, such as which variables most contribute to a prediction or whether some observations are less well fit. Interpretability becomes more difficult when the model is nonlinear. Nonlinear models occur in statistical models with polynomial or interaction terms between quantitative predictors, and almost all computational models such as random forests, support vector machines, or neural networks (e.g. Breiman, 2001a; Boser et al., 1992; Anderson, 1995).

In linear models interpretation of the importance of variables is relatively straightforward, one adjusts for the covariance of multiple variables when examining the relationship with the response. The interpretation is valid for the full domain of the predictors. In nonlinear models, one needs to consider the model in small neighborhoods of the domain to make any assessment of variable importance. Even though this is difficult, it is especially important to interpret model fits as we become more dependent on nonlinear models for routine aspects of life to avoid issues described in Stahl (2021). Understanding how nonlinear models behave when usage extrapolates outside the domain of predictors, either in sub-spaces where few samples were provided in the training set, or extending outside the domain. It is especially important because nonlinear models can vary wildly and predictions can be dramatically wrong in these areas.

Explainable Artificial Intelligence (XAI) is an emerging field of research focused on methods for the interpreting of models (Adadi and Berrada, 2018; Barredo Arrieta et al., 2020). A class of techniques, called *local explanations* (LEs), provide methods to approximate linear variable importance, called local variable attributions (LVAs), at the location of each observation or the predictions at a specific point in the data domain. Because these are point-specific, it is challenging to comprehensively visualize them to understand a model. There are common approaches for visualizing high-dimensional data as a whole, but what is needed are new approaches for viewing these individual LVAs relative to the whole.

For multivariate data visualization, a *tour* (Asimov, 1985; Buja and Asimov, 1986; Lee et al., 2021) of linear data projections onto a lower-dimensional space, could be an element of XAI, complementing LVAs. Applying tours to model interpretation is recommended by Wickham et al. (2015) primarily to examine the fitted model in the space of the data. Cook et al. (2007) describe the use of tours for exploring classification boundaries and model diagnostics (Caragea et al., 2008; Lee et al., 2013; da Silva et al., 2021). There are various types of tours. In a *manual* or radial tour (Cook and Buja, 1997; Spyrison and Cook, 2020), the path of linear projections is defined by changing the contribution of a selected variable. We propose to use this to scrutinize the LVAs. This approach could be considered to be a counter-factual, what-if analysis, such as *ceteris paribus* ("other things held constant") profiles (Biecek, 2020).

The remainder of this paper is organized as follows. Section 2 covers the background of the LEs and the traditional visuals produced. Section 3 explains the tours and particularly the radial manual tour. Section 4 discusses the visual layout in the graphical user interface and how it facilitates analysis, data pre-processing, and package infrastructure. Illustrations are provided in Section 5 for a range of supervised learning tasks with categorical and quantitative response variables. These show how the LVAs can be used to get an overview of the model's use of predictors and to investigate errors in the model predictions. Section 6 concludes with a summary of the insights gained. The methods are implemented in the **R** package **cheem**.

2 Local Explanations

LVAs shed light on machine learning model fits by estimating linear variable importance in the vicinity of a single observation. There are many approaches for calculating LVAs. A comprehensive summary of the taxonomy of currently available methods is provided in Figure 6 by Barredo Arrieta et al. (2020). It includes a large number of model-specific explanations such as deepLIFT (Shrikumar et al., 2016, 2017), a popular recursive method for estimating importance in neural networks. There are fewer model-agnostic methods, of which LIME (Ribeiro et al., 2016) and SHaply Additive exPlanations (SHAP) (Lundberg and Lee, 2017), are popular.

These observation-level explanations are used in various ways depending on the data. In image classification, where pixels correspond to predictors, saliency maps overlay or offset a heatmap to indicate important pixels (Simonyan et al., 2014). For example, pixels corresponding to snow may be highlighted as important contributors when distinguishing if a picture contains a coyote or husky. In text analysis, word-level contextual sentiment analysis highlights the sentiment and magnitude of influential words (Vanni et al., 2018). In the case of numeric regression, they are used to explain additive contributions of variables from the model intercept to the observation's prediction (Ribeiro et al., 2016).

We will be focusing on SHAP values in this paper, but the approach is applicable to any method used to calculate the LVAs. SHAP calculates the variable contributions of one observation by examining the effect of other variables on the predictions. The term "SHAP" refers to Shapley (1953)'s method to evaluate an individual's contribution in cooperative games by assessing this player's performance in the presence or absence of other players. Strumbelj and Kononenko (2010) introduced SHAP for LEs in machine learning models. Variable importance can depend on the sequence in which variables are entered into the model fitting process, thus for any sequence we get a set of variable contribution values for a single observation. These values will add up to the difference between the fitted value for the observation, and the average fitted value for all observations. Using all possible sequences, or permutations, gives multiple values for each variable, which are averaged to get the SHAP value for an observation. It can be helpful to standardize variables prior to computing SHAP values if they have been measured on different scales.

The approach is related to partial dependence plots (for example see chapter 8 of Molnar (2022)), used to explain the effect of a variable by predicting the response for

a range of values on this variable after fixing the value of all other variables to their mean. Though partial dependence plots are a global approximation of the variable importance, while SHAP is specific to one observation.

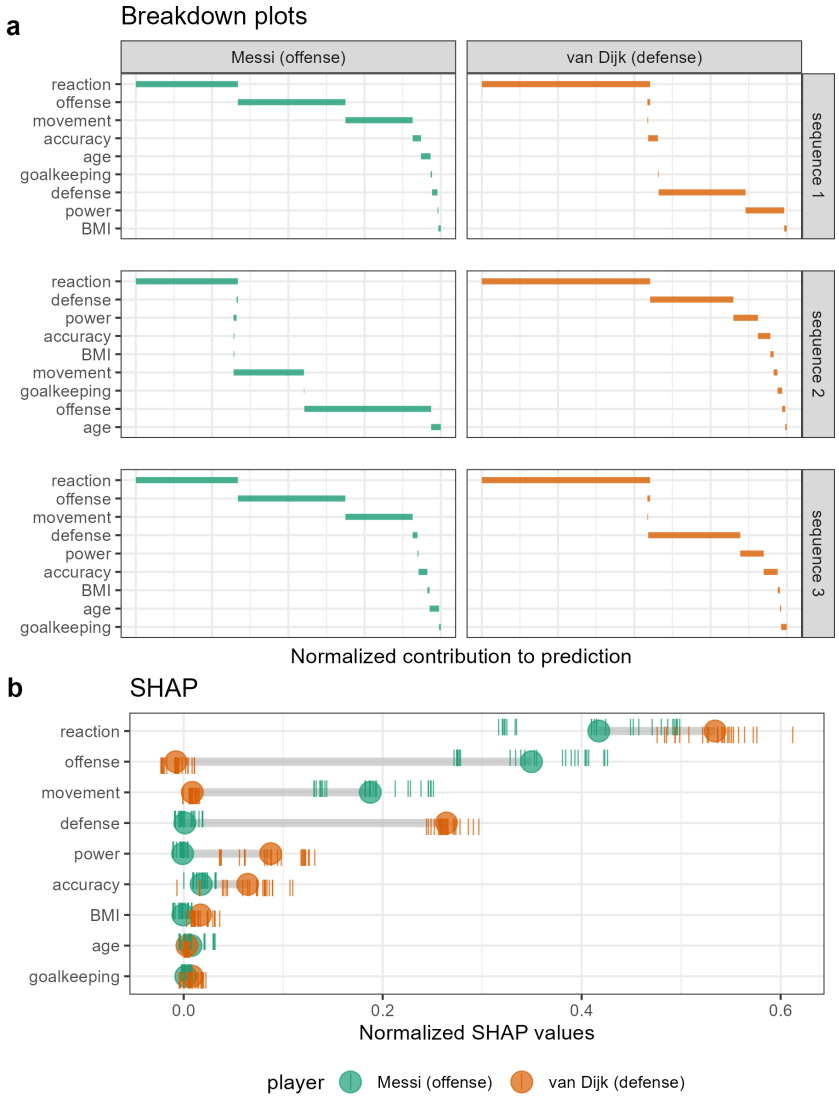


Fig. 1 Illustration of SHAP values for a random forest model FIFA 2020 player wages from nine skill predictors. A star offensive and defensive player are compared, L. Messi and V. van Dijk, respectively. Panel (a) shows breakdown plots of three sequences of the variables. The sequence of the variables impacts the magnitude of their attribution. Panel (b) shows the distribution of attribution for each variable across 25 sequences of predictors, with the mean displayed as a dot for each player. Reaction skills are important for both players. Offense and movement are important for Messi but not van Dijk, and conversely, defense and power are important for van Dijk but not Messi.

We use 2020 season FIFA data (Leone, 2020) to illustrate SHAP following the procedures described in Biecek and Burzykowski (2021). There are 5000 observations of nine predictor variables measuring players' skills and one response variable, wages (in euros). A random forest model is fit regressing players' wages on the skill variables. In this illustration in Figure 1 the SHAP values are compared for a star offensive player (L. Messi) and a prominent defensive player (V. van Dijk). We are interested in knowing how the skill variables locally contribute to the wage prediction of each player. A difference in the attribution of the variable importance across the two positions of the players can be expected. This would be interpreted as how a player's salary depends on which combination of skills. Panel (a) is a version of a breakdown plot (Gosiewska and Biecek, 2019) where just three sequences of variables are shown, for two observations. A breakdown plot shows the absolute values of the variable attribution for an observation, usually sorted from the highest value to the lowest. There is no scale on the horizontal axis here because values are considered relative to each other. Here we can see how the variable contribution can change depending on sequence, relative to both players. (Note that the order of the variables is different in each plot because they have been sorted by the biggest average contribution across both players.) For all sequences, and for both players `reaction` has the strongest contribution, with perhaps more importance for the defensive player. Then it differs by player: for Messi `offense` and `movement` have the strongest contributions, and for van Dijk it is `defense` and `power`, regardless of the variable sequence.

Panel (b) shows the differences in the player's median values (large dots) for 25 such sequences (tick marks). We can see that the wage predictions for the two players come from different combinations of skill sets, as might be expected for players whose value on the team depends on their offensive or defensive prowess. It is also interesting to see from the distribution of values across the different sequences of variables, that there is some multimodality. For example, look at the SHAP values for `reaction` for Messi, and in some sequences, reaction has a much lower contribution than others. This suggests that other variables (`offense`, `movement` probably) can substitute for `reaction` in the wage prediction.

This can also be considered similar to examining the coefficients from all subsets regression, as described in Wickham et al. (2015). Various models that are similarly good might use different combinations of the variables. Examining the coefficients from multiple models helps to understand the relative importance of each variable in the context of all other variables. This is similar to the approach here with SHAP values, that by examining the variation in values across different permutations of variables, we can gain more understanding of the relationship between the response and predictors.

For the application, we use *tree SHAP*, a variant of SHAP that enjoys a lower computational complexity (Lundberg et al., 2018). Instead of aggregating over sequences of the variables, tree SHAP calculates observation-level variable importance by exploring the structure of the decision trees. Tree SHAP is only compatible with tree-based models. so random forests are used for illustration.

There are numerous R packages currently available on CRAN that provide functions for computing SHAP and other LVA values, including `treeshap` (Kominsarczyk et al., 2023), `fastshap` (Greenwell, 2023), `kernelshap` (Mayer and Watson, 2023),

`shapr` (Sellereite et al., 2023), `shapviz` (Mayer, 2023b), `PPtreeregViz` (Lee and Cho, 2022), `ExplainPrediction` (Robnik-Sikonja, 2018), `flashlight` (Mayer, 2023a), and the package `DALEX` has many resources (Biecek, 2018). Molnar (2022) provides good explanations of the different methods and how to apply them to different models.

3 Tours and the Radial Tour

A *tour* enables the viewing of high-dimensional data by animating many linear projections with small incremental changes. It is achieved by following a path of linear projections (bases) of high-dimensional space. One key variable of the tour is the object permanence of the data points; one can track the relative change of observations in time and gain information about the relationships between points across multiple variables. There are various types of tours that are distinguished by how the paths are generated (Lee et al., 2021; Cook et al., 2008).

The manual tour (Cook and Buja, 1997) defines its path by changing a selected variable's contribution to a basis to allow the variable to contribute more or less to the projection. The requirement constrains the contribution of all other variables that a basis needs to be orthonormal (columns correspond to vectors, with unit length, and orthogonal to each other). The manual tour is primarily used to assess the importance of a variable to the structure visible in a projection. It also lends itself to pre-computation queued in advance or computed on the fly for human-in-the-loop analysis (Karwowski, 2006).

A version of the manual tour called a *radial tour* is implemented in Spyrison and Cook (2020) and forms the basis of this new work. In a radial tour, the selected variable can change its magnitude of contribution but not its angle; it must move along the direction of its original contribution. The implementation allows for pre-computation and interactive re-calculation to focus on a different variable. In this work, the radial tour allows us to explore the sensitivity of LVA to the prediction of a model.

4 The Cheem Viewer

To explore the LVAs, coordinated views (Roberts, 2007) (also known as ensemble graphics, Unwin and Valero-Mora, 2018) are provided in the *cheem viewer* application. There are two primary plots: the **global view** to give the context of all of the SHAP values and the **radial tour view** to explore the LVAs with user-controlled rotation. There are numerous user inputs, including variable selection for the radial tour and observation selection for making comparisons. There are different plots used for the categorical and quantitative responses. Figures 3 and 4 are screenshots showing the cheem viewer for the two primary tasks: classification (categorical response) and regression (quantitative response).

4.1 Global View

The global view provides context for all observations and facilitates the exploration of the separability of the data and attribution spaces. The attribution space refers to

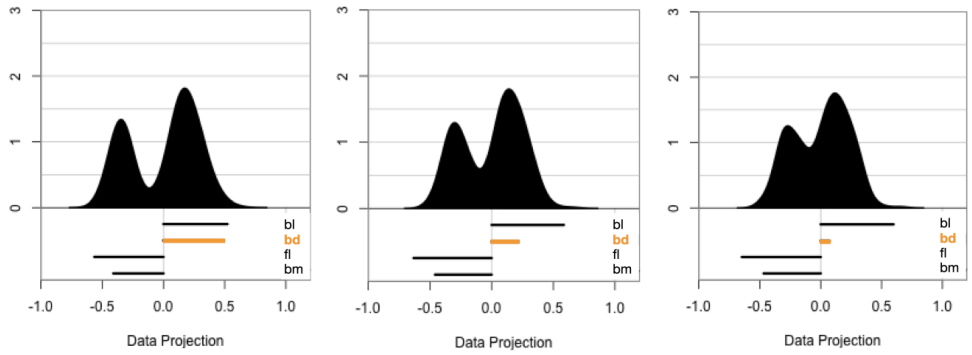


Fig. 2 The radial tour allows the user to remove a variable from a projection, to examine the importance of this variable to the structure in the plot. Here we have a 1D projection of the penguins data displayed as a density plot. The line segments on the bottom correspond to the coefficients of the variables making up the projection. The structure in the plot is bimodality (left), and the importance of the variable `bd` is being explored. As this variable contribution is reduced in the plot (middle, right) we can see that the bimodality decreases. Thus `bd` is an important variable contributing to the bimodal structure.

the SHAP values for each observation. These spaces both have dimensionality $n \times p$, where n is the number of observations and p is the number of variables.

The visualization is composed of the first two principal components of the data (left) and the attribution (middle) spaces. These single 2D projections will not reveal all of the structure of higher-dimensional space, but they are helpful visual summaries. In addition, a plot of the observed against predicted response values is also provided (Figures 3c, 4b) to help identify observations poorly predicted by the model. For classification tasks, color indicates the predicted class and misclassified observations are circled in red. Linked brushing between the plots is provided (click and drag), and a tabular display of selected points helps to facilitate the exploration of the spaces and the model (shown in Figure 4d).

While the comparison of these spaces is interesting, the primary purpose of the global view is to enable the selection of particular observations to explore in detail. We have designed it to enable a comparison between an observation that is interesting in some way, perhaps misclassified, or poorly predicted, relative to an observation with similar predictor values but a more expected prediction. For brevity, we call the interesting observation the primary investigation (PI), and the other is the comparison investigation (CI). These observations are highlighted as an asterisk and $\times$, respectively.

4.2 Radial Tour

The radial tour is used to explore how the SHAP value of a variable relates to it's effect on the predicted value. In a similar way as explained in Section 3, where the radial tour is used to understand a variable's contribution to cluster structure, for model prediction explanations, the radial tour is used to understand a variable's contribution

to the observation's predicted value. By altering the contribution using the radial tour, we see how the predicted value might change. If a small change in the variable contribution results in a big change in predicted value, then this variable substantially explains the model prediction. The SHAP values are estimates of the local importance, and provide a good starting place from which to begin a radial tour. They can be misleading, and the radial tour can help to assess the strength of the explanatory power of the SHAP value. Because the SHAP values are local, using linear projections to explore a local neighborhood of a nonlinear model is reasonable.

There are two plots in this part of the interface. The first (Figures 3e and 4e) is a display of the SHAP values for all observations. This will generally give the global view of variables important for the fit as a whole, but it will also highlight observations that have different patterns. The second plot is the radial tour, which for classification is a density plot of a 1D projection (Figure 3f), and for regression are scatterplots of the observed response values, and residuals, against a 1D projection (Figure 4f).

The LVAs for all observations are normalized (sum of squares equals 1), and thus, the relative importance of variables can be compared across all observations. These are depicted as a vertical parallel coordinate plot (Ocagne, 1885). (The SHAP values of the PI and CI are shown as dashed and dotted lines, respectively.) One should obtain a sense of the overall importance of variables from this plot. The more important variables will have larger values, and in the case of classification tasks variables that have different magnitudes for different classes are more globally important. For example, Figure 3e suggests that `bl` is important for distinguishing the green class from the other two. For regression, one might generally observe which variables have low values for all observations (not important). For example, `BMI` and `pwr` in Figure 4e, have a range of high and low values (e.g., `off`, `def`), suggesting they are important for some observations and not important for others.

A bar chart is overlaid to represent the projection shown in the radial tour on the right. It starts from the SHAP values of the PI, but if the user changes the projection the length of these bars will reflect this change. By scaling the SHAP value it becomes an (attribution) projection.

The attribution projection of the PI is the initial 1D basis in a radial tour, displayed as a density plot for a categorical response (Figure 3f) and as scatterplots for a quantitative response (Figure 4f). The PI and CI are indicated by vertical dashed and dotted lines, respectively. The radial tour varies the contribution of the selected variable. This is viewed as an animation of the projections from many intermediate bases. Doing so tests the sensitivity of structure (class separation or strength of relationship) to the variable's contribution. The CI attribution of the CI does not impact the bases but it highlighted from context. For classification, if the separation between classes diminishes when the variable contribution is reduced, this suggests that the variable is important for class separation. For regression, if the relationship scatterplot weakens when the variable contribution is reduced, indicating that the variable is important for accurately predicting the response.

The purpose of using both the PI and CI when using the radial tour is comparison. Remember the CI is a representative individual with an expected prediction (correct class or small residual) and the PI is a particularly interesting individual with

a less expected prediction. The radial tour would start from the attribution projection corresponding to the SHAP values of the PI, and vary the contribution of a variable where the SHAP values differ from those of the CI. The goal is then to examine how the model prediction would change for the PI if the variable contribution changed, to be more similar to that of the CI.

4.3 Classification Task

Selecting a misclassified observation as PI and a correctly classified point nearby in data space as CI makes it easier to examine the variables most responsible for the error. The global view (Figure 3c) displays the model confusion matrix. The radial tour is 1D and displays as density where color indicates class. An animation slider enables users to vary the contribution of variables to explore the sensitivity of the separation to that variable.

4.4 Regression Task

Selecting an inaccurately predicted observation as PI and an accurately predicted observation with similar variable values as CI is a helpful way to understand how the model is failing or not. The global view (Figure 4a) shows a scatterplot of the observed vs predicted values, which should exhibit a strong relationship if the model is a good fit. The points can be colored by a statistic, residual, a measure of outlyingness (log Mahalanobis distance), or correlation to aid in understanding the structure identified in these spaces.

In the radial tour view, the observed response and the residuals (vertical) are plotted against the attribution projection of the PI (horizontal). The attribution projection can be interpreted similarly to the predicted value from the global view plot. It represents a linear combination of the variables, and a good fit would be indicated when there is a strong relationship with the observed values. This can be viewed as a local linear approximation if the fitted model is nonlinear. As the contribution of a variable is varied, if the value of the PI does not change much, it would indicate that the prediction for this observation is NOT sensitive to that variable. Conversely, if the predicted value varies substantially, the prediction is very sensitive to that variable, suggesting that the variable is very important for the PI's prediction.

4.5 Interactive Variables

The application has several reactive inputs that affect the data used, aesthetic display, and tour manipulation. These reactive inputs make the software flexible and extensible (Figure 3a & d). The application also has more exploratory interactions to help link points across displays, reveal structures found in different spaces, and access the original data.

A tooltip displays the observation number/name and classification information while the cursor hovers over a point. Linked brushing allows the selection of points (left click and drag) where those points will be highlighted across plots (Figure 4a & b). The information corresponding to the selected points is populated on a dynamic

Fig. 3 Overview of the cheem viewer for classification tasks (categorical response). Global view inputs, (a), set the PI, CI, and color statistic. Global view, (b) PC1 by PC2 approximations of the data- and attribution-space. (c) prediction by observed y (visual of the confusion matrix for classification tasks). Points are colored by predicted class, and red circles indicate misclassified observations. Radial tour inputs (d) select variables to include and which variable is changed in the tour. (e) shows a parallel coordinate display of the distribution of the variable attributions while bars depict contribution for the current basis. The black bar is the variable being changed in the radial tour. Panel (f) is the resulting data projection indicated as density in the classification case.

table (Figure 4d). These interactions aid the exploration of the spaces and, finally, the identification of primary and comparison observations.

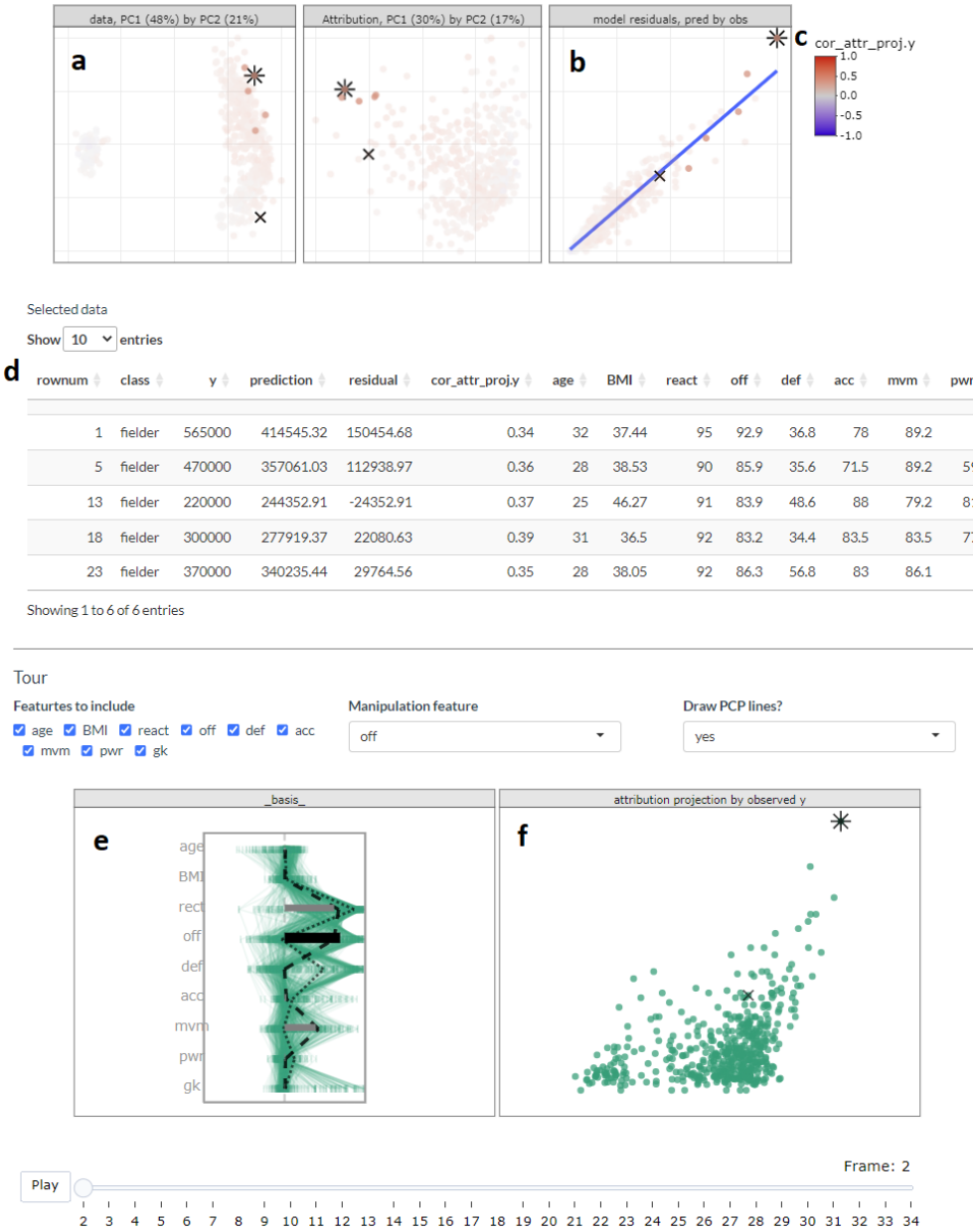


Fig. 4 Overview of the cheem viewer for regression tasks (quantitative response) and illustration of interactive variables. Panel (a) PCA of the data- and attributions- spaces and the (b) observed vs predicted values. Four selected points are highlighted in the PC spaces and tabularly displayed. Coloring on a statistic (c) highlights the structure organized in the attribution space. Interactive tabular display (d) populates when observations are selected. Contribution of the 1D basis affecting the horizontal position (e) parallel coordinate display of the variable attribution from all observations, and horizontal bars show the contribution to the current basis. Regression projection (f) uses the same horizontal projection and fixes the vertical positions to the observed y and residuals (middle and right).

4.6 Preprocessing

It is vital to mitigate the render time of visuals, especially when users may want to iterate many explorations. All computational operations should be prepared before run time. The work remaining when an application is run solely reacts to inputs and rendering visuals and tables. Below discusses the steps and details of the reprocessing.

- **Data:** predictors and response are unscaled complete numerical matrix. Most models and local explanations are scale-invariant. Keep the normality assumptions of the model in mind.
- **Model:** any model and compatible explanation could be explored with this method. Currently, random forest models are applied via the package **randomForest** (Liaw and Wiener, 2002), compatibility tree SHAP. Modest hyperparameters are used. Namely, classification models use 125 trees, number of variables at each split (mtry) of $\sqrt{p}$, and minimum terminal node size of $max(1, n/500)$. While regression models use 125 tree, $p/3$ variables at split, and $max(5, n/500)$ minimum terminal node size.
- **Local explanation:** Tree SHAP is calculated for *each* observation using the package **treeshap** (Kominsarczyk et al., 2023). We opt to find the attribution of each observation in the training data and not fit to fit variable interactions.
- **Cheem viewer:** after the model and full explanation space are calculated, each variable is scaled by standard deviations away from the mean to achieve common support for visuals. Statistics for mapping to color are computed on the scaled spaces.

The time to preprocess the data will vary significantly with the complexity of the model and the LE. For reference, the FIFA data contained 5000 observations of nine explanatory variables that took 2.5 seconds to fit a random forest model of modest hyperparameters. Extracting the tree SHAP values of each observation took 270 seconds in total. PCA and statistics of the variables and attributions took 2.8 seconds. These run times were from a non-parallelized session on a modern laptop, but suffice it to say that most of the time will be spent on the LVA. An increase in model complexity or data dimensionality will quickly become an obstacle. Its reduced computational complexity makes tree SHAP an excellent candidate to start. Alternatively, some package and methods use approximate calculations of LEs, such as **fastshap** Greenwell (2020).

5 Case Studies

To illustrate the cheem method it is applied to modern data sets, two classification examples and then two of regression.

5.1 Palmer Penguins, Species Classification

The Palmer penguins data (Gorman et al., 2014; Horst et al., 2020) was collected on three species of penguins foraging near Palmer Station, Antarctica. The data is publicly available to substitute for the overly-used iris data and is quite similar in form. After removing incomplete observations, there are 333 observations of four physical

measurements, bill length (`bl`), bill depth (`bd`), flipper length (`fl`), and body mass (`bm`) for this illustration. A random forest model was fit with species as the response variable.

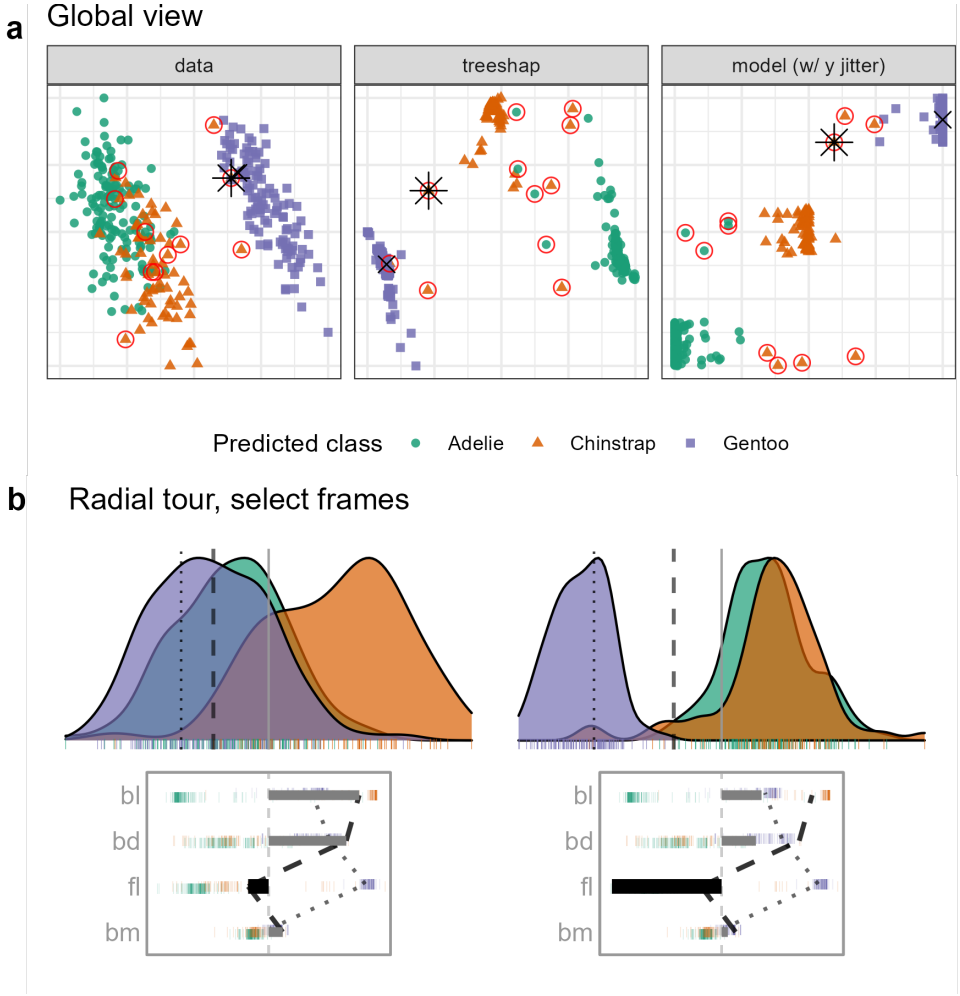


Fig. 5 Examining the SHAP values for a random forest model classifying Palmer penguin species. The PI is a Gentoo (purple) penguin that is misclassified as a Chinstrap (orange), marked as an asterisk in (a) and the dashed vertical line in (b). The radial view shows varying the contribution of 'fl' from the initial attribution projection (b, left), which produces a linear combination where the PI is more probably (higher density value) a Chinstrap than a Gentoo (b, right). (The animation of the radial tour is at https://vimeo.com/666431172.)

Figure 5 shows plots from the cheem viewer for exploring the random forest model on the penguins data. Panel (a) shows the global view, and panel (b) shows several 1D projections generated with the radial tour. Penguin 243, a Gentoo (purple), is the PI because it has been misclassified as a Chinstrap (orange).

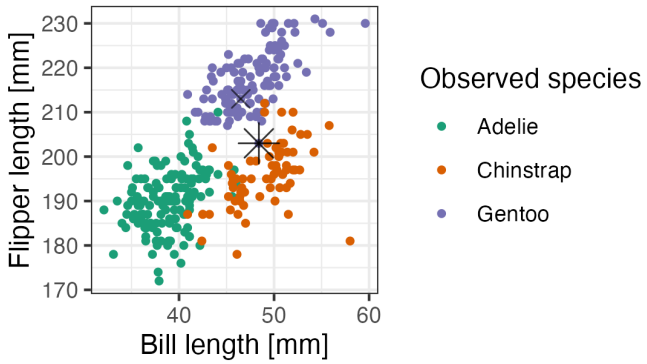


Fig. 6 Checking what is learned from the cheem viewer. This is a plot of flipper length ('fl') and bill length ('bl'), where an asterisk highlights the PI. A Gentoo (purple) misclassified as a Chinstrap (orange). The PI has an unusually small 'fl' length which is why it is confused with a Chinstrap.

There is more separation visible in the attribution space than in the data space, as would be expected. The predicted vs observed plot reveals a handful of misclassified observations. A Gentoo which has been wrongly labeled as a Chinstrap is selected for illustration. The PI is a misclassified point (represented by the asterisk in the global view and a dashed vertical line in the tour view). The CI is a correctly classified point (represented by an $\times$ and a vertical dotted line).

The radial tour is used here to examine which variable most contributed to the incorrect classification of the PI, to understand why the model was prediction differed from that of the CI. It starts from the attribution projection of the misclassified observation (b, left). The important variables identified by SHAP in the (wrong) prediction for this observation are mostly `bl` and `bd` with small contributions of `fl` and `bm`. This projection is a view where the Gentoo (purple) looks much more likely for this observation than Chinstrap. That is, this combination of variables is not particularly useful because the PI looks very much like other Gentoo penguins. The radial tour is used to vary the contribution of flipper length (`fl`) to explore this. (In our exploration, this was the third variable explored. It is typically helpful to explore the variables with more significant contributions, here `bl` and `bd`. Still, when doing this, nothing was revealed about how the PI differed from other Gentoos). On varying `fl`, as it contributes increasingly to the projection (b, right), more and more, this penguin looks like a Chinstrap. This suggests that `fl` should be considered an important variable for explaining the (wrong) prediction.

Figure 6 confirms that flipper length (`fl`) is vital for the confusion of the PI as a Chinstrap. Here, flipper length and body length are plotted, and the PI can be seen to be closer to the Chinstrap group in these two variables, mainly because it has an unusually low value of flipper length relative to other Gentoos. From this view, it makes sense that it is a hard observation to account for, as decision trees can only partition only vertical and horizontal lines.

5.2 Chocolates, Milk/Dark Classification

The chocolates data set consists of 88 observations of ten nutritional measurements determined from their labels and labeled as either milk or dark. Dark chocolate is considered healthier than milk. Students collected the data during the Iowa State University class STAT503 from nutritional information on the manufacturer's websites and were normalized to 100g equivalents. The data is available in the **cheem** package. A random forest model is used for the classification of chocolate types.

It could be interesting to examine the nutritional properties of any dark chocolates that have been misclassified as milk. A reason to do this is that a dark chocolate, nutritionally more like milk should not be considered a healthy alternative. It is interesting to explore which nutritional variables contribute most to the misclassification.

This type of exploration is shown in Figure 7, where a chocolate labeled dark but predicted to be milk is chosen as the PI (observation 22). It is compared with a CI that is a correctly classified dark chocolate (observation 7). The PCA plot and the tree SHAP PCA plots (a) show a big difference between the two chocolate types but with confusion for a handful of observations. The misclassifications are more apparent in the observed vs predicted plot and can be seen to be mistaken in both ways: milk to dark and dark to milk.

The attribution projection for chocolate 22 suggests that Fiber, Sugars, and Calories are most responsible for its incorrect prediction. The way to read this plot is to see that Fiber has a large negative value while Sugars and Calories have reasonably large positive values. In the density plot, observations on the very left of the display would have high values of Fiber (matching the negative projection coefficient) and low values of Sugars and Calories. The opposite would be interpreting a point with high values in this plot. The dark chocolates (orange) are primarily on the left, and this is a reason why they are considered to be healthier: high fiber and low sugar. The density of milk chocolates is further to the right, indicating that they generally have low fiber and high sugar.

The PI (dashed line) can be viewed against the CI (dotted line). Now, one needs to pay attention to the parallel coordinate plot of the SHAP values, which are local to a particular observation, and the density plot, which is the same projection of all observations as specified by the SHAP values of the PI. The variable contribution of the two different predictions can be quickly compared in the parallel coordinate plot. The PI differs from the comparison primarily on the Fiber variable, which suggests that this is the reason for the incorrect prediction.

From the density plot, which is the attribution projection corresponding to the PI, both observations are more like dark chocolates. Using the radial tour to vary the contribution of Sugars, results in it being removed and replaced by Fiber, and reason for the wrong classification becomes apparent. In this 1D projection observation 22 is more similar to milk chocolates, suggests that Fiber is the culprit for the model mistakenly seeing it as a milk chocolate.

It would also be interesting to explore an inverse misclassification, where a milk chocolate is misclassified as a dark chocolate. Chocolate 84 is selected and is compared with a correctly predicted milk chocolate (observation 71). The corresponding global view and radial tour frames are shown in Figure 8.

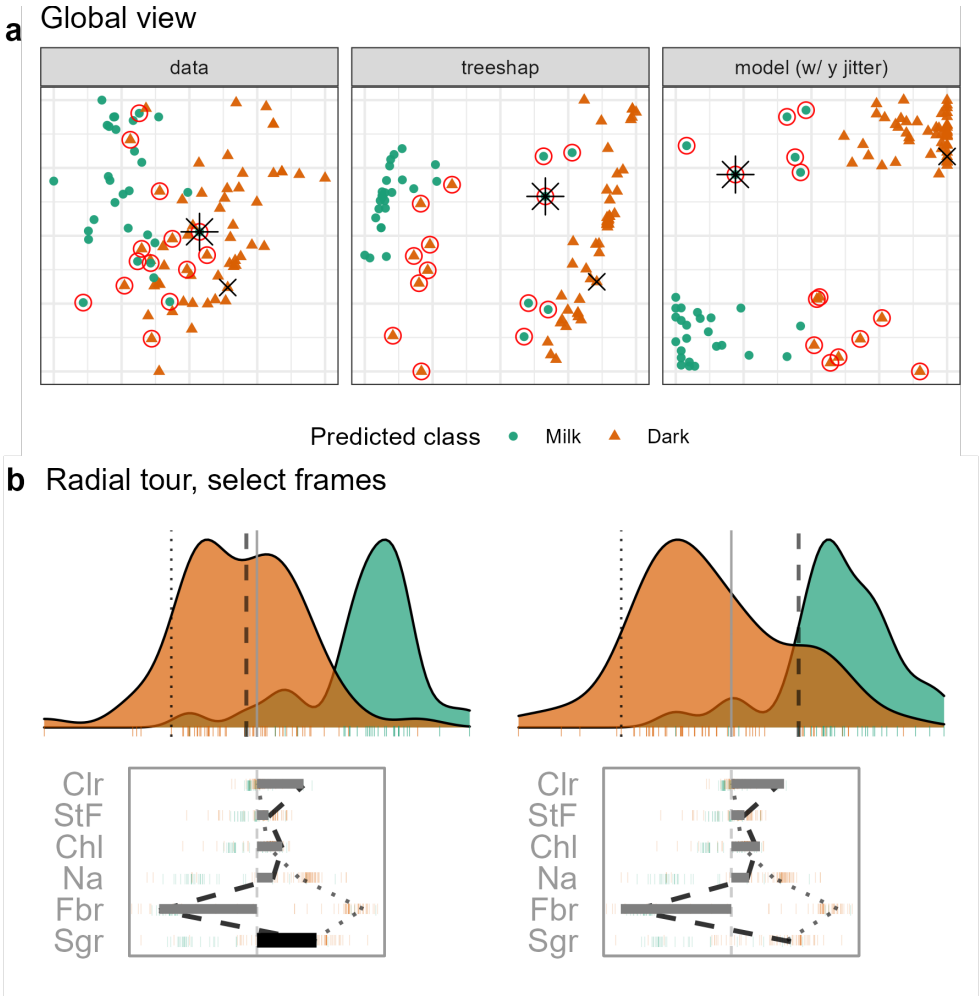


Fig. 7 Examining the LVA for a PI which is dark (orange) chocolate incorrectly predicted to be milk (green). From the attribution projection, this chocolate correctly looks more like dark than milk, which suggests that the LVA does not help understand the prediction for this observation. So, the contribution of Sugar is varied—reducing it corresponds primarily with an increased magnitude from Fiber. When Sugar is zero, Fiber contributes strongly toward the left. In this view, the PI is closer to the bulk of the milk chocolates, suggesting that the prediction put a lot of importance on Fiber. This chocolate is a rare dark chocolate without any Fiber leading to it being mistaken for a milk chocolate. (A video of the tour animation can be found at https://vimeo.com/666431143.)

Comparing the attributions of the PI and the CI, large differences in the values of Sodium and Fiber can be seen. The contribution of Sodium is selected to be varied in the radial tour. From the density plot of the initial attribution projection, the PI is equally likely to be milk or dark dark chocolate. When the contribution of Sodium is increased, the balance shifts, and the PI is more likely to be correctly considered to be a milk chocolate. This supports that the model prediction was erroneous because it didn't adequately consider the value of Sodium in making the prediction.

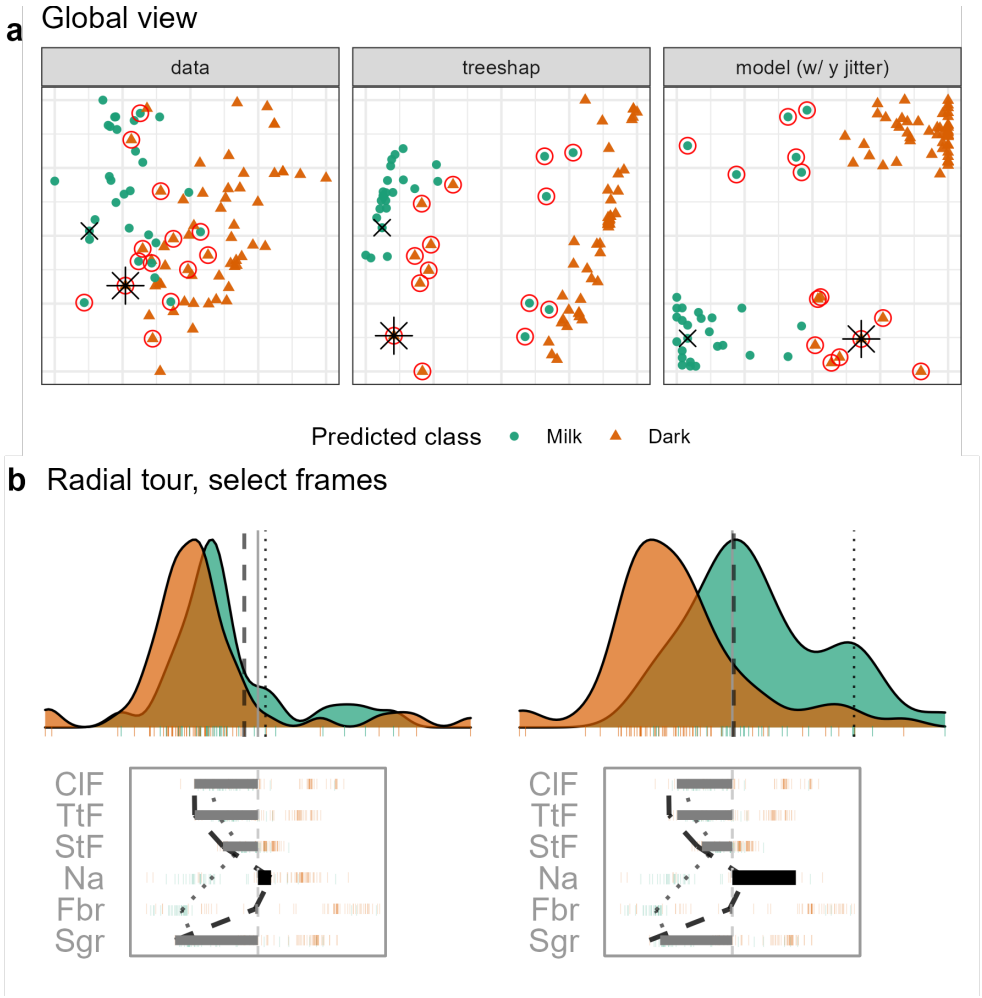


Fig. 8 Examining the LVA for a PI which is a milk (green) chocolate incorrectly predicted to be a dark (orange). From the density plot of the attribution projection the PI could equally likely be milk or dark, where as the CI is more definitely milk. Sodium and Fiber have the largest differences in attributed variable importance, with values lose to zero, instead of large negative values like other milk chocolates. The lack of importance attributed to these variables is suspected of contributing to the mistake. When the contribution of Sodium is changedr, we see if the model had used a larger contribution of Sodium to make the prediction the PI would have likely been predicted to be a milk chocolate. (A video of the tour animation can be found at https://vimeo.com/666431148.)

5.3 FIFA, Wage Regression

The 2020 season FIFA data (Leone, 2020; Biecek, 2018) contains many skill measurements of soccer/football players and wage information. Nine higher-level skill groupings were identified and aggregated from highly correlated variables. A random forest model is fit from these predictors, regressing on player wages [2020 euros]. The

model was fit from 5000 observations before being thinned to 500 players to mitigate occlusion and render time. Continuing from the information in Section 2, we are interested to see the difference in attribution based on what is known about different players, that is a leading offensive fielder (L. Messi) as compared with a top defensive fielder (V. van Dijk). (These same observations were shown in Figure 1.) With the radial tour we can explore how these players wages might be predicted if their skill sets were different.

Figure 9 tests the support of the LVA for the PI (Messi). The contribution from `def` is varied in the radial tour, in contrast to offensive skills (`off`). As the contribution of defensive skills increases, Messi's wage plummets. Messi's predicted wage would be much lower defensive skills played a larger role in the prediction - the model prediction reinforces that he is clearly not getting paid for his ability to defend.

Although we don't show it here, offensive and reaction (`rct`) skills are both crucial to explaining the star offensive player's predicted wage. If the contribution of either is changed, the other substitutes! That is, rotating one variable out, results in the other rotating in, when the radial tour is used, and the wage value does not change, remaining in a far-right location in the plot. Some change in predicted wage is seen if instead the contribution of a variable with low importance is varied.

5.4 Ames Housing, Sales Price Regression

Ames housing data (De Cock, 2011) was subset to North Ames, with 338 house sales. A random forest model was fit, predicting the sale price [USD] from the property variables: Lot Area (`LtA`), Overall Quality (`Qlt`), Year the house was Built (`YrB`), Living Area (`LvA`), number of Bathrooms (`Bth`), number of Bedrooms (`Bdr`), the total number of Rooms (`Rms`), Year the Garage was Built (`GYB`), and Garage Area (`GrA`). Using interactions with the global view, a house with an extreme negative residual and an accurate observation with a similar prediction is selected.

Figure 10 illustrates the exploration of the model predictions for the house sale 74 (PI), which is under-valued by the model. The CI has a similar predicted price though the prediction was accurate. The SHAP values for the PI and CI have very different values of Lot Area. The attribution projection would give the PI a higher value than the CI, suggesting that the Lot Area value is important for the predicted value of the PI but not for that of the CI. As the contribution of Lot Area is decreased in the radial tour, the predict value of PI decreases while the CI increases. This is quite interesting, that the SHAP value picks up the importance of Lot Area. And it appears that the model does not adequately use this variable. For the attribution projection, with a large contribution from Lot Area, the PI is better predicted than in the model, and would have a smaller residual.

6 Discussion

There is a clear need to provide more tools interpret black box models. Techniques such as SHAP, LIME, Break-down calculate LEs for each observation in the data. They estimate how important variables are for the model's prediction of a single observation.

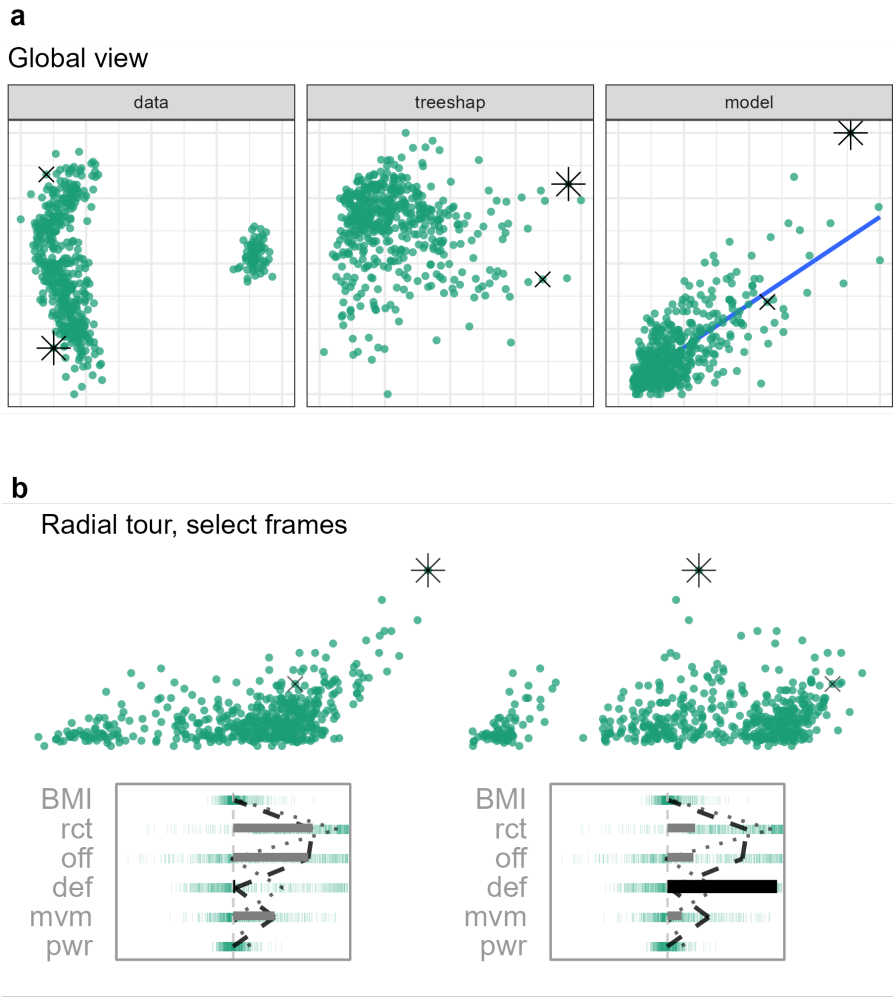


Fig. 9 Exploring the wages relative to skill measurements in the FIFA 2020 data. Star offensive player (L. Messi) is the PI, and he is compared with a top defensive player (V. van Dijk): (a) global view, (b) observed values vs linear combination of predictors (predicted values). The attribution projection produces a view where Messi has very high predicted (and observed) wages. Defense ('def') is the chosen variable to vary. It starts very low, and Messi's predicted wages decrease dramatically as its contribution increases (right plot). The increased contribution in defense comes at the expense of offensive and reaction skills. The interpretation is that Messi's high wages are most attributable to his offensive and reaction skills, as initially provided by the LVA. (A video of the animated radial tour can be found at https://vimeo.com/666431163.)

This paper has provided additional interactive graphics tools to utilize LEs to explore and understand model predictions. Several diagnostic plots are provided to assist with understanding the sensitivity of a prediction to particular variables. A global view shows the data space, explanation space, and residual plot, to get an

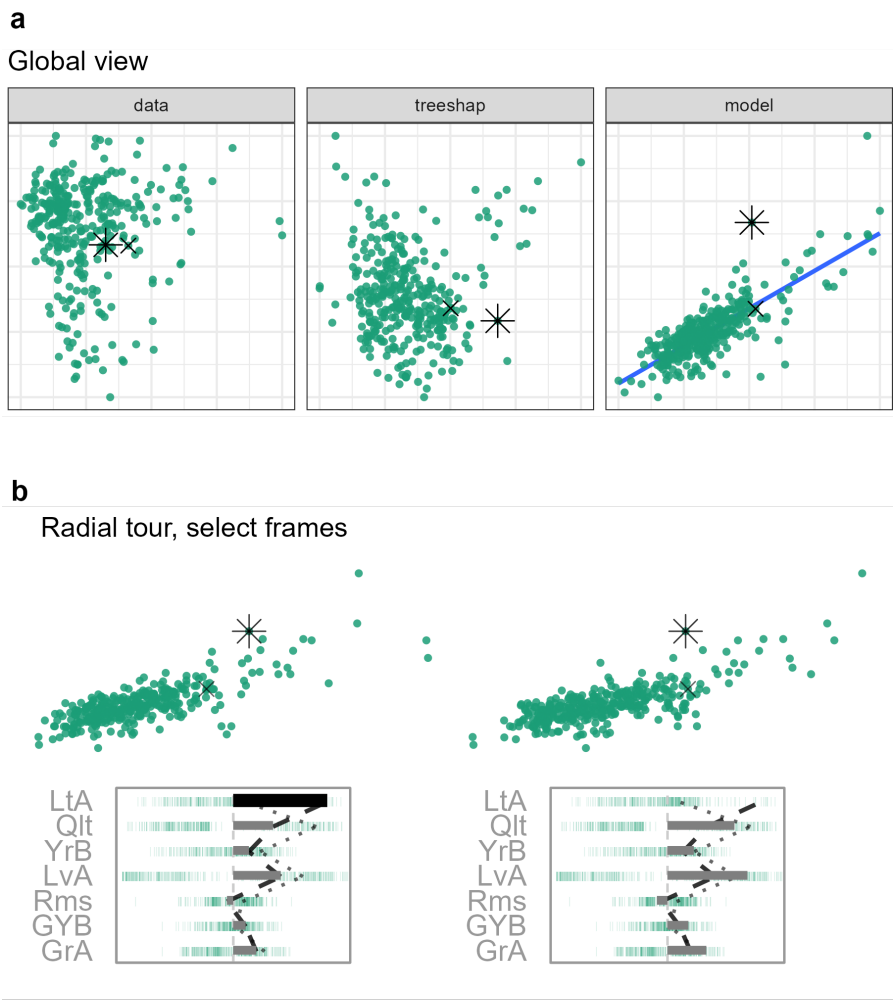


Fig. 10 Exploring an observation with an extreme residual as the PI in relation to an observation with an accurate prediction for a similarly priced house in a random forest fit to the Ames housing data: (a) global view, (b) observed values vs linear combination of predictors (predicted values). The LVA indicates a sizable attribution to Lot Area (*LtA*), while the CI has minimal attribution to this variable. The PI has a higher predicted value than the CI in the attribution projection. Reducing the contribution of Lot Area brings these two prices in line. This suggests that if the model did not value Lot Area so highly for this observation, then the observed sales price would be quite similar. That is, the large residual in the model is due to a lack of factoring Lot Area into the prediction of PI's sales price. (A video showing the animation is at https://vimeo.com/666431134.)

overview of the distribution of LEs across all observations. The user can interactively select observations to compare, contrast, and study further. The LE is converted into an LVA (linear projection) where the radial tour can be used to understand the prediction's sensitivity to a particular variable.

This approach has been illustrated using four data examples of random forest models with the tree SHAP LVA. LEs focus on the model fit and help to dissect which variables are most responsible for the fitted value. They can also form the basis of learning how the model has got it wrong, when the observation is misclassified or has a large residual.

In the penguins example, we showed how the misclassification of a penguin arose due to it having an unusually small flipper size compared to others of its species. This was verified by making a follow-up plot of the data. The chocolates example shows how a dark chocolate was misclassified primarily due to its attribution to Fiber, and a milk chocolate was misclassified as dark due to its lowish Sodium value. In the FIFA example, we show how low Messi's salary would be if it depended on their defensive skill. In the Ames housing data, an inaccurate prediction for a house was likely due to the lot area not being effectively used by the random forest model.

This analysis is manually intensive and thus only feasible for investigating a few observations. The recommended approach is to investigate an observation where the model has not predicted accurately and compare it with an observation with similar predictor values where the model fitted well. The radial tour launches from the attribution projection to enable exploration of the sensitivity of the prediction to any variable. It can be helpful to make additional plots of the variables and responses to cross-check interpretations made from the cheem viewer. This methodology provides an additional tool in the box for studying model fitting.

These tools work better for smaller data, because being able to interact with the plots is necessary. XAI has been developed to tackle large data. To work with bigger data sets, would involve subsetting it after modeling and computing the LEs, to keep a representative sample of well-fitted observations, along with the observations that are especially interesting to investigate.

There are many additional future directions for this work. Primarily, development should make it easier to focus on what can be learned from the LEs, to be able to compare different versions, to flag or annotate values, and output of log the results of interactive analysis.

7 Package Infrastructure

An implementation is provided in the open-source **R** package **cheem**, available on CRAN Spyrison (2023). Example data sets are provided. You can upload your own data after model fitting and computing the LVAs. The LVAs need to be pre-computed, possibly using the `cheem_ls()` function, and saved as an `rds` file. Examples show how to do this for tree SHAP values, using **treeshap** (tree-based models from **gbm**, **lightgbm**, **randomForest**, **ranger**, or **xgboost** Greenwell et al. (2020); Shi et al. (2022); Liaw and Wiener (2002); Wright and Ziegler (2017); Chen et al. (2021), respectively). The SHAP and oscillation explanations could be easily added using `DALEX::explain()` (Biecek, 2018; Biecek and Burzykowski, 2021).

The application was made with **shiny** (Chang et al., 2021). The tour visual is built with **spinifex** (Spyrison and Cook, 2020). Both views are created first with **ggplot2** (Wickham, 2016) and then rendered as interactive `html` widgets with **plotly**

(Sievert, 2020). **DALEX** (Biecek, 2018) and *Explanatory Model Analysis* (Biecek and Burzykowski, 2021) are helpful for understanding LEs and how to apply them.

The package can be installed from CRAN, and the application can be run using the following **R** code:

```
install.packages("cheem", dependencies = TRUE)
library("cheem")
run_app()
```

A version of the cheem application can be accessed at https://nicholas-spyrison.shinyapps.io/cheem/, the development version of the package is available at https://github.com/nspyrison/cheem, and documentation of the package can be found at https://nspyrison.github.io/cheem/.

Acknowledgments. Kim Marriott provided advice on many aspects of this work, especially on the explanations in the applications section. This research was supported by the Australian Government Research Training Program (RTP) scholarships. Thanks to Jieyang Chong for helping proofread this article. The namesake, Cheem, refers to a fictional race of humanoid trees from Doctor Who lore. **DALEX** pulls on from that universe, and we initially apply tree SHAP explanations specific to tree-based models.

References

Adadi, A. and M. Berrada. 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access* 6: 52138–52160 .

Anderson, J.A. 1995. *An introduction to neural networks*. Cambridge: MIT press.

Asimov, D. 1985. The Grand Tour: a Tool for Viewing Multidimensional Data. *SIAM journal on scientific and statistical computing 6*(1): 128–143. https://doi.org/10.1137/0906011 .

Barredo Arrieta, A., N. Diaz-Rodriguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera. 2020. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion* 58: 82–115. https://doi.org/ https://doi.org/10.1016/j.inffus.2019.12.012 .

Biecek, P. 2018. DALEX: explainers for complex predictive models in R. *The Journal of Machine Learning Research 19*(1): 3245–3249 .

Biecek, P. 2020. *ceterisParibus: Ceteris Paribus Profiles*.

Biecek, P. and T. Burzykowski. 2021. *Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models*. New York: CRC Press.

Boser, B.E., I.M. Guyon, and V.N. Vapnik 1992. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational*

learning theory, pp. 144–152.

Breiman, L. 2001a. Random forests. *Machine learning 45*(1): 5–32 .

Breiman, L. 2001b. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science 16*(3): 199–231 .

Buja, A. and D. Asimov 1986. Grand Tour Methods: An Outline. In *Proceedings of the Seventeenth Symposium on the Interface of Computer Sciences and Statistics on Computer Science and Statistics*, New York, NY, USA, pp. 63–67. Elsevier North-Holland, Inc.

Caragea, D., D. Cook, H. Wickham, and V. Honavar 2008. *Visual Methods for Examining SVM Classifiers*, pp. 136–153. Berlin, Heidelberg: Springer Berlin Heidelberg.

Chang, W., J. Cheng, J. Allaire, C. Sievert, B. Schloerke, Y. Xie, J. Allen, J. McPherson, A. Dipert, and B. Borges. 2021. *shiny: Web Application Framework for R.*

Chen, T., T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou, M. Li, J. Xie, M. Lin, Y. Geng, and Y. Li. 2021. xgboost: Extreme Gradient Boosting.

Cook, D. and A. Buja. 1997. Manual Controls for High-Dimensional Data Projections. *Journal of Computational and Graphical Statistics 6*(4): 464–480. https://doi.org/10.2307/1390747 .

Cook, D., A. Buja, E.K. Lee, and H. Wickham. 2008. Grand Tours, Projection Pursuit Guided Tours, and Manual Controls, *Handbook of Data Visualization*, 295–314. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-33037-0_13.

Cook, D., D.F. Swayne, and A. Buja. 2007, December. *Interactive and Dynamic Graphics for Data Analysis: With R and GGobi*. Springer Science & Business Media.

da Silva, N., D. Cook, and E.K. Lee. 2021. A Projection Pursuit Forest Algorithm for Supervised Classification. *Journal of Computational and Graphical Statistics*: 1–21 .

De Cock, D. 2011. Ames, Iowa: Alternative to the Boston housing data as an end of semester regression project. *Journal of Statistics Education 19*(3) .

Gorman, K.B., T.D. Williams, and W.R. Fraser. 2014. Ecological sexual dimorphism and environmental variability within a community of Antarctic penguins (genus Pygoscelis). *PloS one 9*(3): e90081 .

Gosiewska, A. and P. Biecek. 2019. IBreakDown: Uncertainty of model explanations for non-additive predictive models. *arXiv preprint arXiv:1903.11420* .

Greenwell, B. 2020. *fastshap: Fast Approximate Shapley Values*.

Greenwell, B. 2023. *fastshap: Fast Approximate Shapley Values*. R package version 0.1.0.

Greenwell, B., B. Boehmke, J. Cunningham, and G.B.M. Developers. 2020. gbm: Generalized Boosted Regression Models.

Horst, A.M., A.P. Hill, and K.B. Gorman. 2020. palmerpenguins: Palmer Archipelago (Antarctica) penguin data.

Karwowski, W. 2006. *International Encyclopedia of Ergonomics and Human Factors, -3 Volume Set*. CRC Press.

Kominsarczyk, K., P. Kozminski, S. Maksymiuk, and P. Biecek. 2023, October. *treeshap*.

Lee, E.K. and H. Cho 2022. *PPtreeregViz: Projection Pursuit Regression Tree Visualization*. R package version 2.0.5.

Lee, S., D. Cook, N. da Silva, U. Laa, N. Spyrison, E. Wang, and H.S. Zhang. 2021, December. The state-of-the-art on tours for dynamic visualization of high-dimensional data. *WIREs Computational Statistics n/a*(n/a): e1573. https://doi.org/10.1002/wics.1573 .

Lee, Y., D. Cook, J.w. Park, and E.K. Lee. 2013. PPtree: Projection pursuit classification tree. *Electronic Journal of Statistics* 7: 1369–1386 .

Leone, S. 2020, October. FIFA 20 complete player dataset.

Liaw, A. and M. Wiener. 2002. Classification and regression by randomForest. *R news 2*(3): 18–22 .

Lundberg, S.M., G.G. Erion, and S.I. Lee. 2018. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888* .

Lundberg, S.M. and S.I. Lee 2017. A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems*, pp. 4768–4777.

Mayer, M. 2023a. *flashlight: Shed Light on Black Box Machine Learning Models*. R package version 0.9.0.

Mayer, M. 2023b. *shapviz: SHAP Visualizations*. R package version 0.9.2.

Mayer, M. and D. Watson 2023. *kernelshap: Kernel SHAP*. R package version 0.3.8.

Molnar, C. 2022. *Interpretable Machine Learning* (2 ed.).

Ocagne, M.d. 1885. *Coordonnées parallèles et axiales. Méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèles, par Maurice d'Ocagne, ...* Paris: Gauthier-Villars.

Ribeiro, M.T., S. Singh, and C. Guestrin 2016, August. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, New York, NY, USA, pp. 1135–1144. Association for Computing Machinery.

Roberts, J.C. 2007. State of the art: Coordinated & multiple views in exploratory visualization. In *Fifth international conference on coordinated and multiple views in exploratory visualization (CMV 2007)*, pp. 61–71. IEEE.

Robnik-Sikonja, M. 2018. *ExplainPrediction: Explanation of Predictions for Classification and Regression Models*. R package version 1.3.0.

Sellereite, N., M. Jullum, and A. Redelmeier 2023. *shapr: Prediction Explanation with Dependence-Aware Shapley Values*. R package version 0.2.2.

Shapley, L.S. 1953. *A value for n-person games*. Princeton University Press.

Shi, Y., G. Ke, D. Soukhavong, J. Lamb, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.Y. Liu, and N. Titov. 2022. lightgbm: Light Gradient Boosting Machine.

Shmueli, G. 2010. To explain or to predict? *Statistical science 25*(3): 289–310 .

Shrikumar, A., P. Greenside, and A. Kundaje 2017. Learning important features through propagating activation differences. In *International Conference on Machine Learning*, pp. 3145–3153. PMLR.

Shrikumar, A., P. Greenside, A. Shcherbina, and A. Kundaje. 2016. Not just a black box: Learning important features through propagating activation differences. *arXiv preprint arXiv:1605.01713* .

Sievert, C. 2020. *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC.

Simonyan, K., A. Vedaldi, and A. Zisserman 2014. Deep inside convolutional networks: Visualising image classification models and saliency maps. In *In Workshop at International Conference on Learning Representations*. Citeseer.

Spyrison, N. 2023, November. *cheem: Interactively Explore Local Explanations with the Radial Tour*.

Spyrison, N. and D. Cook. 2020. spinifex: an R Package for Creating a Manual Tour of Low-dimensional Projections of Multivariate Data. *The R Journal 12*(1): 243. https://doi.org/10.32614/RJ-2020-027 .

Stahl, B.C. 2021. Ethical issues of ai. *Artificial Intelligence for a Better Future*: 35–53. https://doi.org/10.1007/978-3-030-69978-9_4 .

Strumbelj, E. and I. Kononenko. 2010. An efficient explanation of individual classifications using game theory. *The Journal of Machine Learning Research* 11: 1–18 .

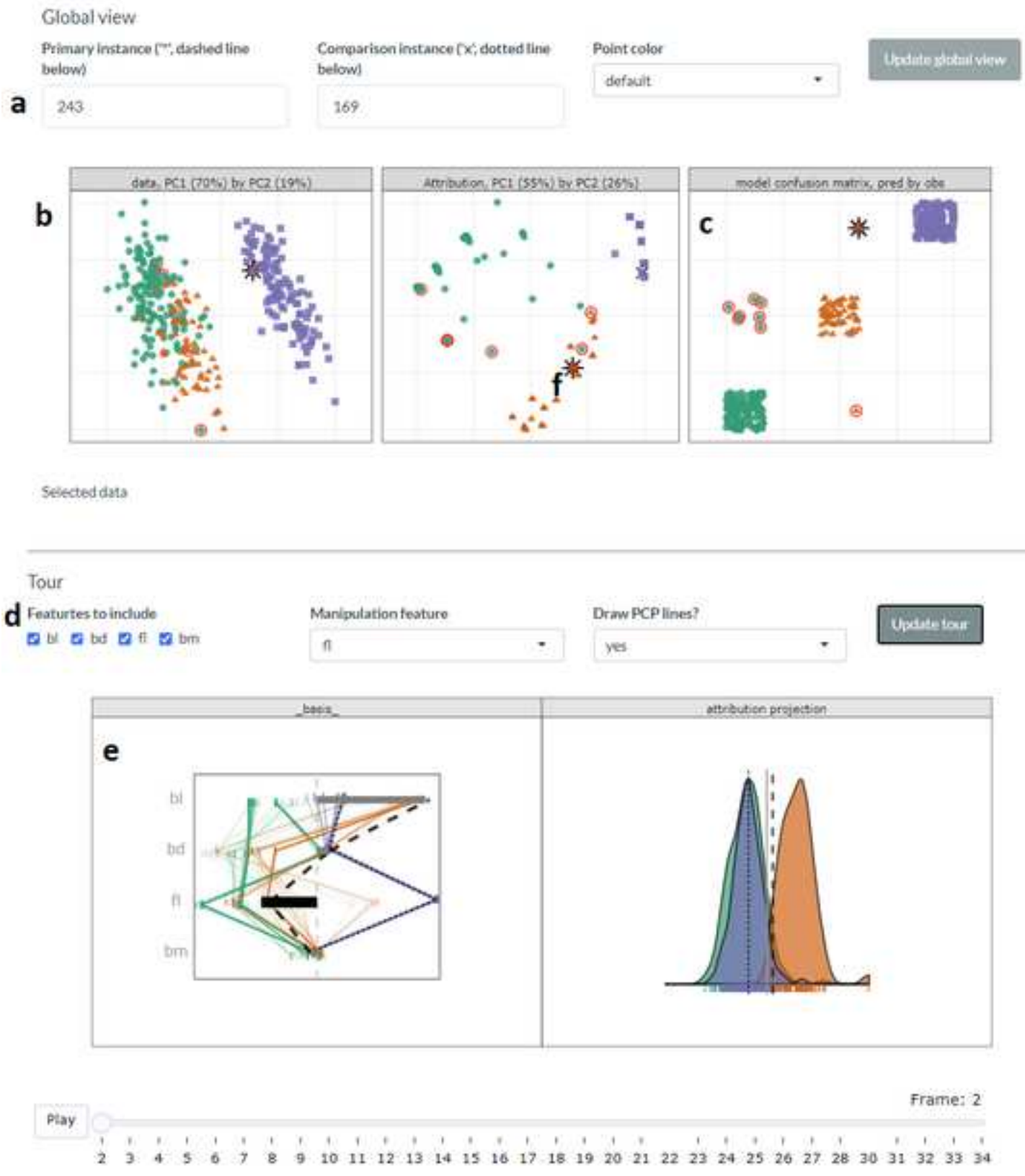
Unwin, A. and P. Valero-Mora. 2018, January. Ensemble Graphics. *Journal of Computational and Graphical Statistics 27*(1): 157–165. https://doi.org/10.1080/10618600.2017.1383264 .

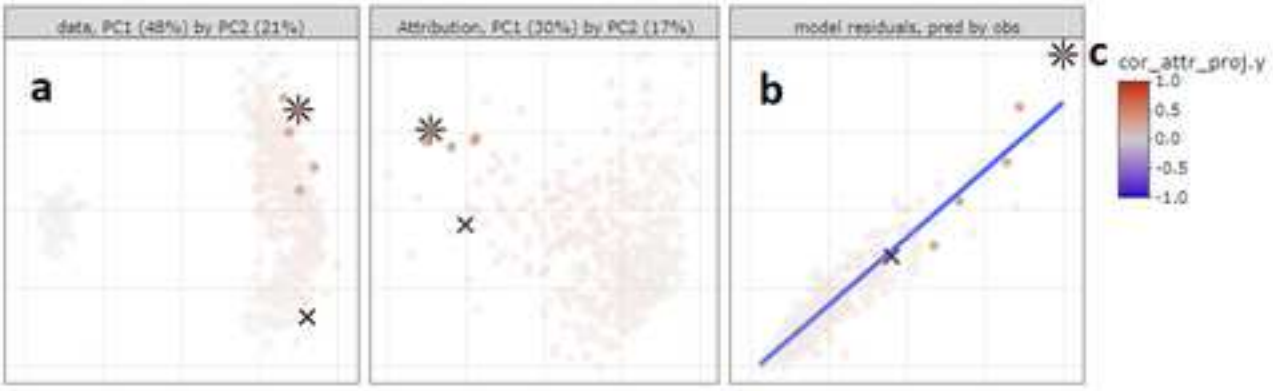
Vanni, L., M. Ducoffe, C. Aguilar, F. Precioso, and D. Mayaffre 2018. Textual Deconvolution Saliency (TDS): a deep tool box for linguistic analysis. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 548–557.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H., D. Cook, and H. Hofmann. 2015, August. Visualizing statistical models: Removing the blindfold. *Statistical Analysis and Data Mining: The ASA Data Science Journal 8*(4): 203–225. https://doi.org/10.1002/sam.11271 .

Wright, M.N. and A. Ziegler. 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software 77*(1): 1–17. https://doi.org/10.18637/jss.v077.i01 .





Selected data

Show 10 entries

d

rownum	class	y	prediction	residual	cor_attr_proj.y	age	BMI	react	off	def	acc	mvm	pwr
1	helder	565000	414545.32	150454.68	0.34	32	37.44	95	92.9	36.8	78	89.2	
5	helder	470000	357061.03	112938.97	0.36	28	38.53	90	85.9	35.6	71.5	89.2	5
13	helder	220000	244352.91	-24352.91	0.37	25	46.27	91	83.9	48.6	88	79.2	8
18	helder	300000	277919.37	22080.63	0.39	31	36.5	92	83.2	34.4	83.5	83.5	7
23	helder	370000	340235.44	29764.56	0.35	28	38.05	92	86.3	56.8	83	86.1	

Showing 1 to 6 of 6 entries

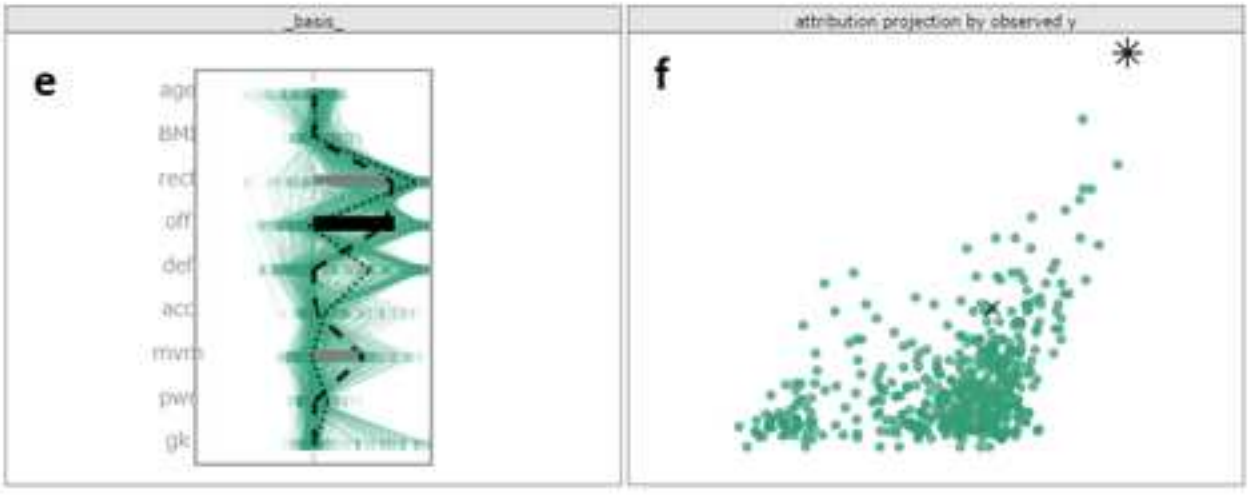
Tour

Featurtes to include

age BMI react off def acc mvm pwr gk

Manipulation feature: off

Draw PCP lines?: yes



Play

Frame: 2

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34