

[Sign up](#)[Sign In](#)

Search Medium



Write



◆ Member-only story

HANDS-ON TUTORIALS

Visualizing Decision Trees with Pybaobabdt

An open-source package for decision tree visualization and model interpretation



Parul Pandey · Follow

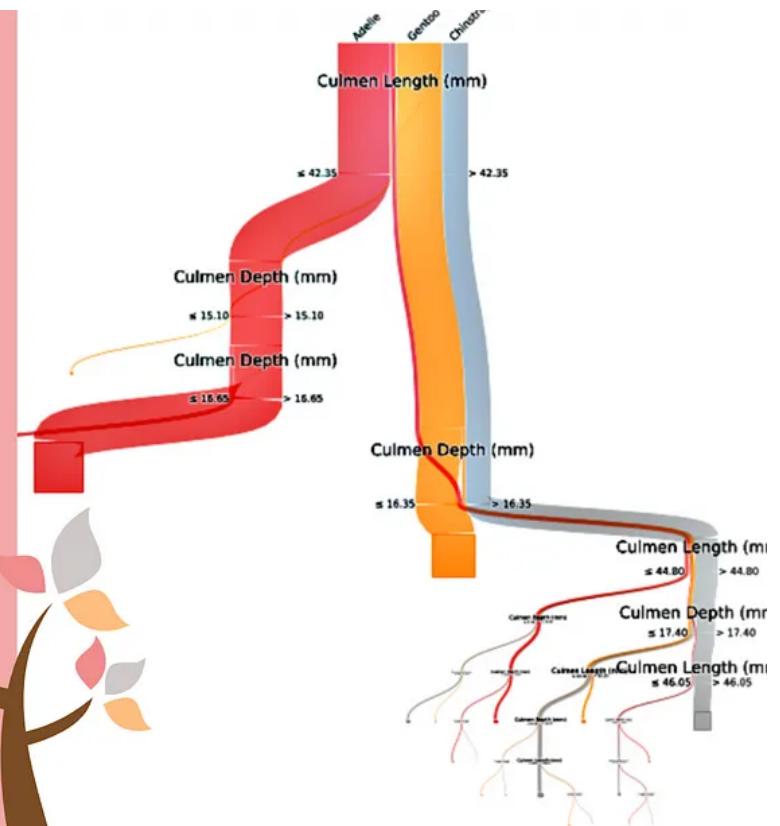
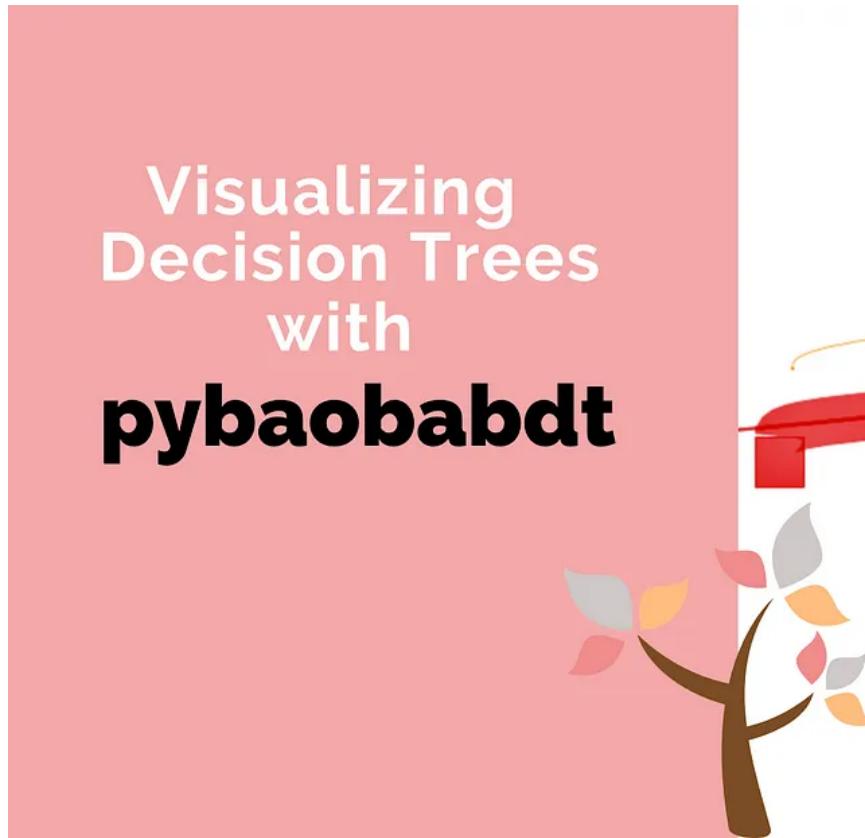
Published in Towards Data Science · 7 min read · Dec 14, 2021

307

6

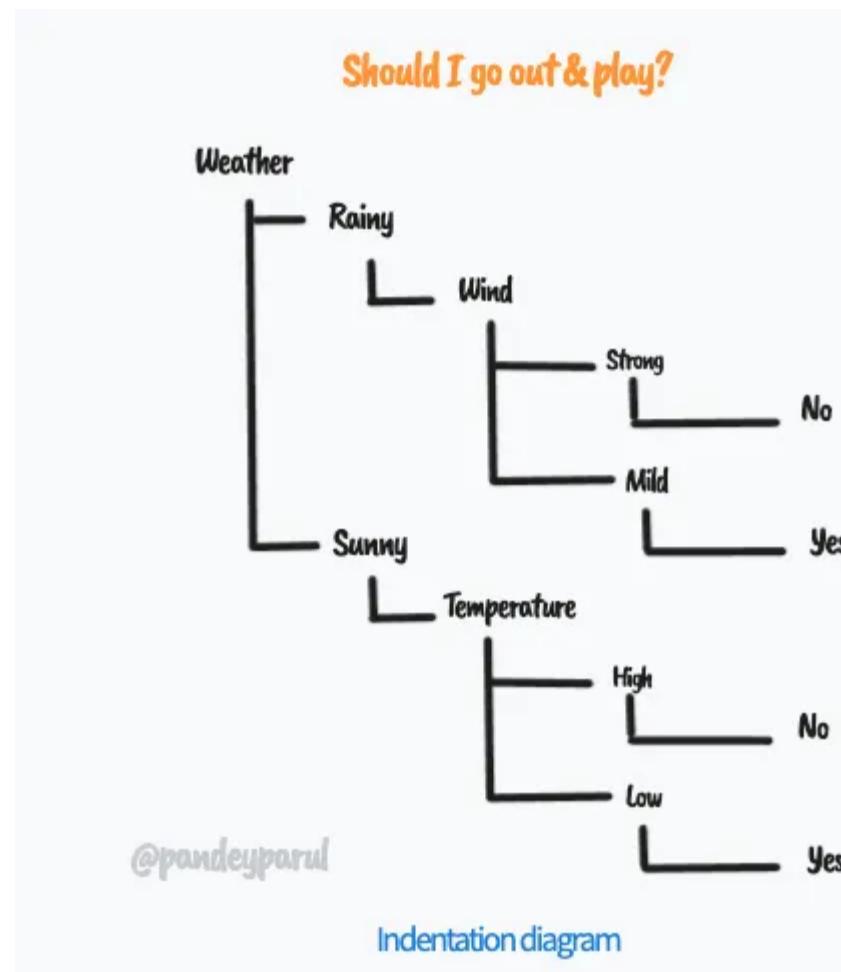
--

6



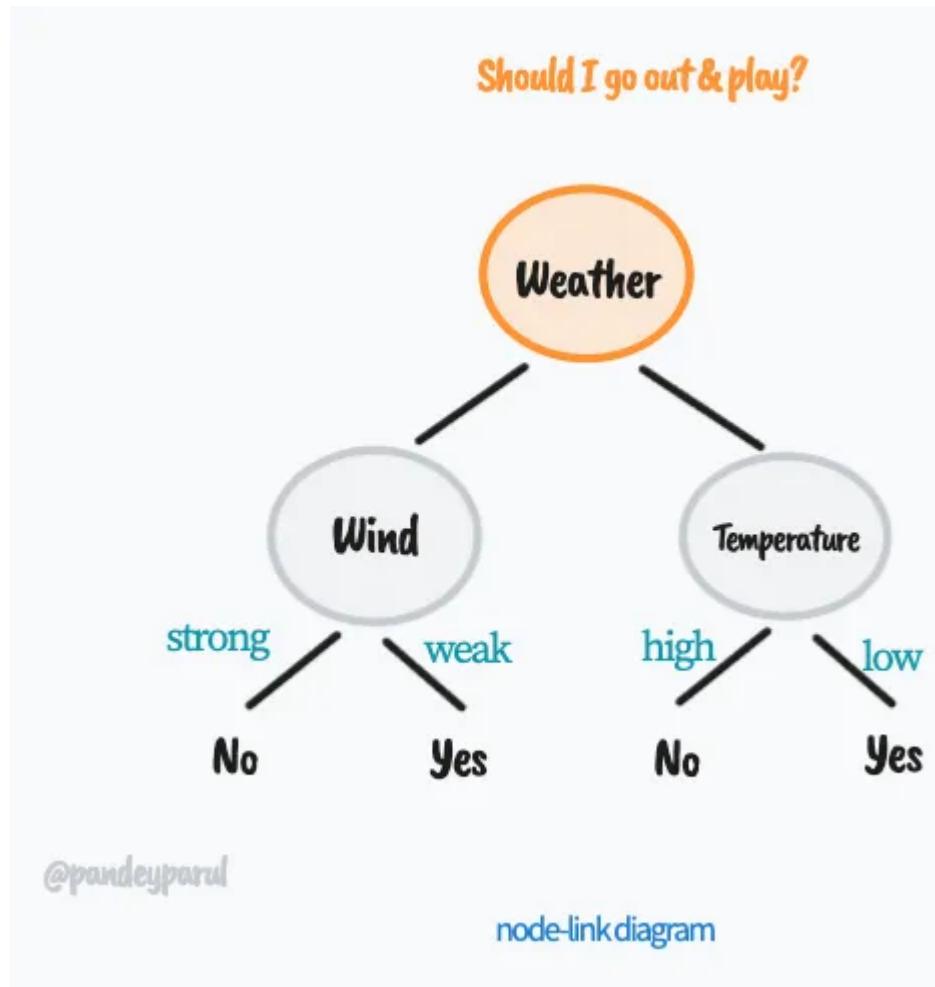
Data visualization is the language of decision-making.
Good charts effectively convey information. Great
charts enable, inform, and improve decision making:
Dante Vitagliano

Decision trees can be visualized in multiple ways. Take, for instance, the **indentation nodes** where every internal and leaf node is depicted as text, while the parent-child relationship is shown by indenting the child with respect to the parent.



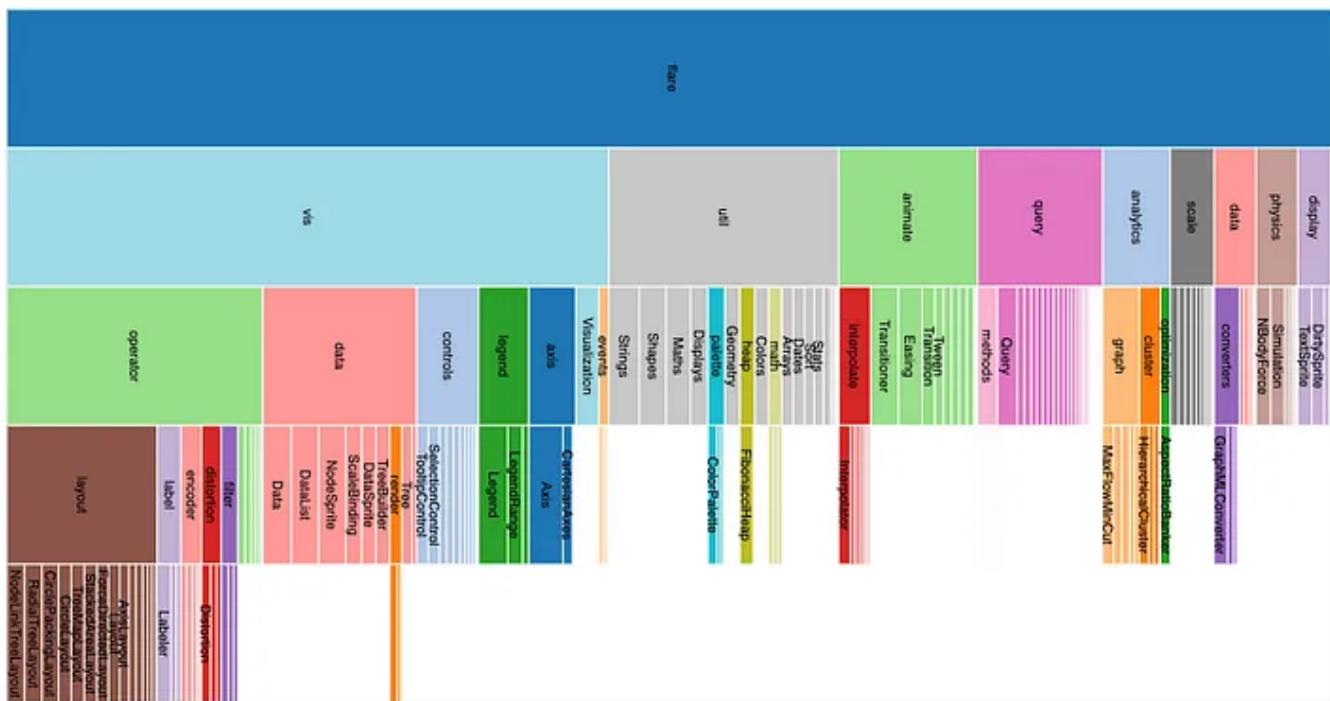
Indentation diagram | Image by Author

Then there is the **node-link diagram**. It is one of the most commonly used methods to visualize decision trees where the nodes are represented via glyphs, and parent and child nodes are connected through links.



node-link diagram| Image by Author

Icicle plots are another option for the same. In addition to displaying the relationship, these plots also help depict the node size. They derive their name from the fact that the resulting visualization looks like icicles.



An icicle plot by

https://www.cs.middlebury.edu/~candrews/showcase/infovis_techniques_s16/icicle_plots/icicleplots.html | CC-BY license

While these techniques are helpful, they do not scale well especially when the size of data increases. In such situations, not only does it become difficult to visualize the data, but interpreting and understanding the tree is also a challenge. **BaobabView** is a library created to overcome such problems, and in this article, we'll look at its python implementation called **pybaobabdt** in detail, along with examples.

A few packages to plot Decision Trees

We initiated the article by discussing the multiple ways of visualizing decision trees. It'll...



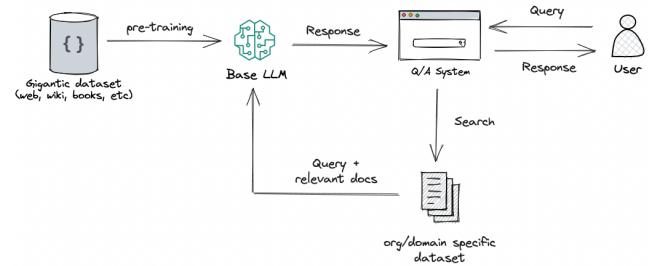
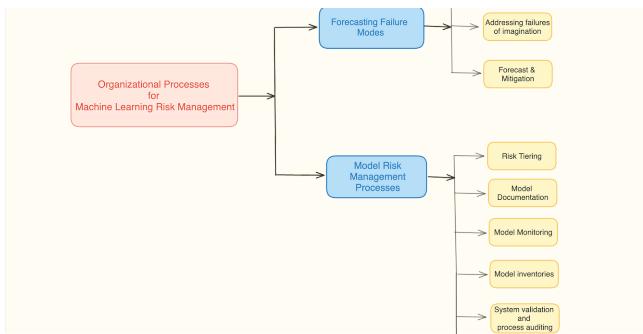
Written by Parul Pandey

21K Followers · Writer for Towards Data Science

[Follow](#)


Principal Data Scientist @H2O.ai | Author of Machine Learning for High-Risk Applications

More from Parul Pandey and Towards Data Science



 Parul Pandey  in Towards Data Science

Organizational Processes for Machine Learning Risk...

Organizational processes are a key nontechnical determinant of reliability in ML...

★ · 7 min read · 5 days ago

 472  2

 +

 2.1K  16

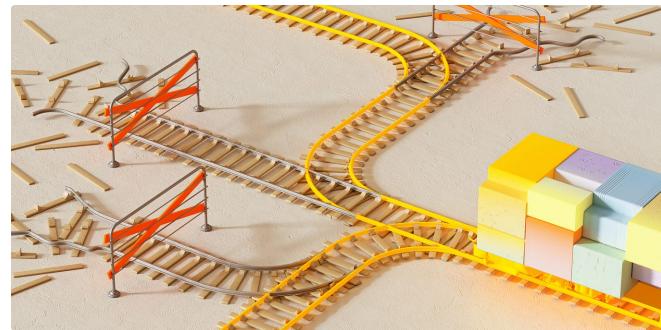
 +

 Heiko Hotz in Towards Data Science

RAG vs Finetuning—Which Is the Best Tool to Boost Your LLM...

The definitive guide for choosing the right method for your use case

★ · 19 min read · Aug 24



Giuseppe Scalamogna in Towards Data Science

New ChatGPT Prompt Engineering Technique: Program Simulation

A potentially novel technique for turning a ChatGPT prompt into a mini-app.

9 min read · Sep 3



1.3K



90


[See all from Parul Pandey](#)
[See all from Towards Data Science](#)

Cultural Competencies for Machine Learning Risk Management

An organization's culture is an essential aspect of responsible AI.

★ · 7 min read · Sep 15



90



Recommended from Medium





Cassie Kozyrkov in Towards Data Science



AshirbadPradhan

Statistics: Are you Bayesian or Frequentist?

The fastest way to diagnose your statistical alignment

6 min read · Jun 4, 2021

4.8K

36



3

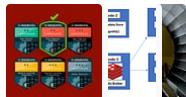


Lists



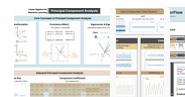
Predictive Modeling w/ Python

20 stories · 428 saves



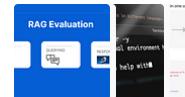
New_Reading_List

174 stories · 122 saves



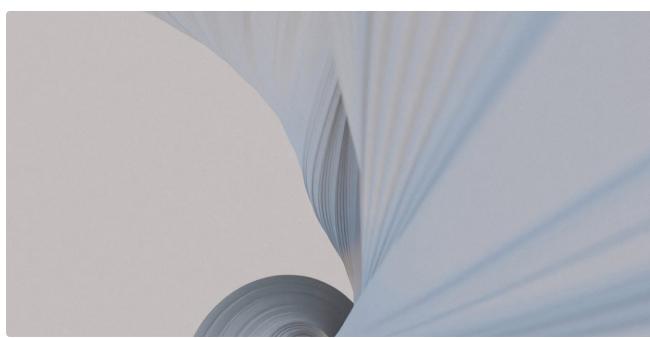
Practical Guides to Machine Learning

10 stories · 491 saves



Natural Language Processing

652 stories · 256 saves

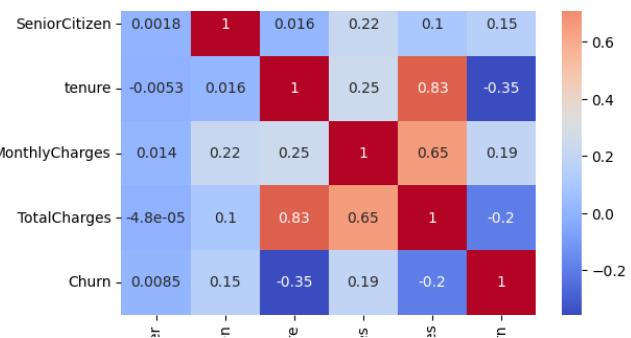


Abel K Widodo

Exploring the Art of Generative AI in Python

Artificial Intelligence (AI) has transformed various industries, and one fascinating area ...

6 min read · Jul 15



Alex Saruni Lodaru

Customer Churn Predictions Using Classification Analysis

Churn refers to the number of customers who stop using a product or service over a given...

13 min read · Jun 4

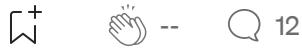


 Prathamesh Gadekar

Seaborn charts that every Data Scientist Knows!

Seaborn charts that every Data Scientist Knows!

7 min read · Apr 3



Google BERT

 Rayyan Shaikh

Mastering BERT: A Comprehensive Guide from Beginner to Advanced...

Introduction: A Guide to Unlocking BERT: From Beginner to Expert

19 min read · Aug 26



[See more recommendations](#)