Understanding Node Splitting in Decision Trees: Exploring GINI, Entropy, and Information Gain

Sai Bharath Bobba, Akarsh Reddy Tatimakula ,Hariharan Kumar

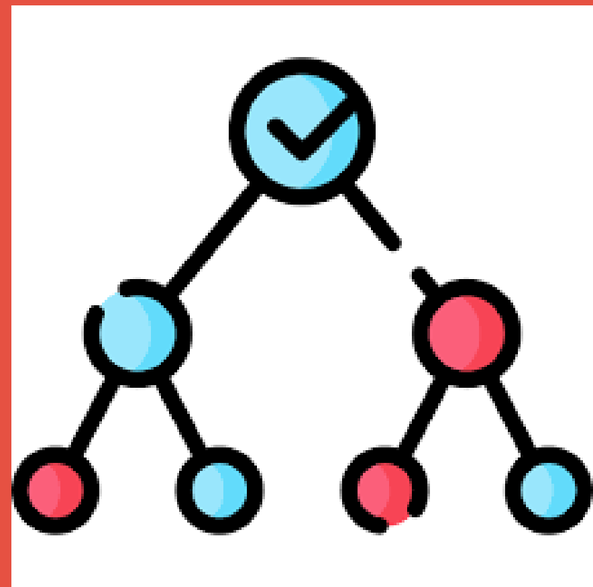
CONTENT

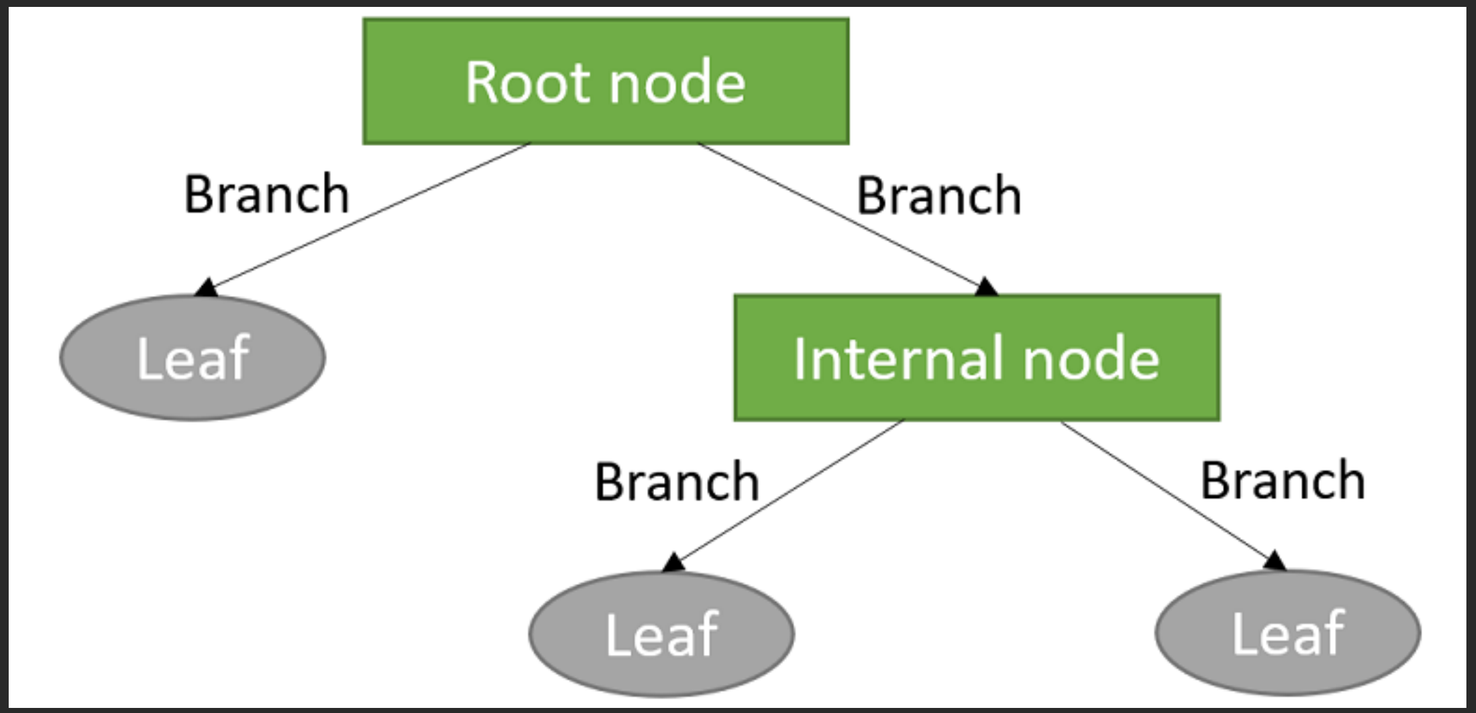
1. Decision Tree
2. Node Splitting
3. Gini
4. Entropy
5. Information Gain
6. Tutorial
7. Conclusion
8. References

02

Decision Tree

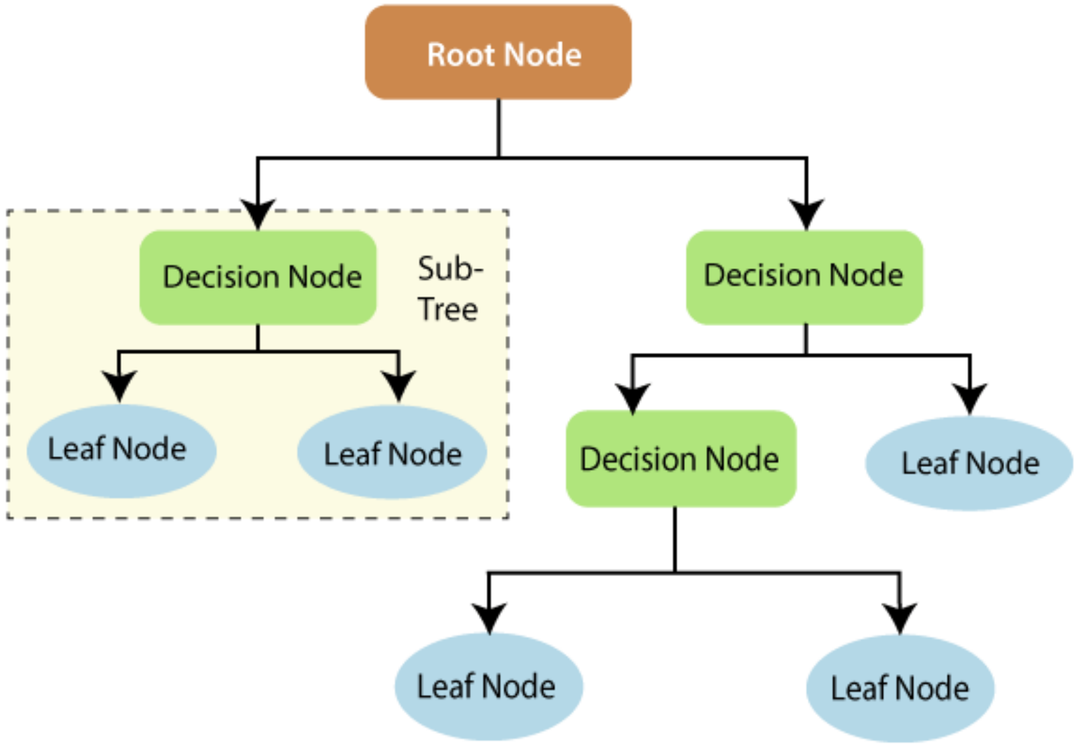
- **A Visual Representation of Choices, Consequences, Probabilities, and Opportunities.**
- **A Way of Breaking Down Complicated Situations Down to Easier-to-Understand Scenarios.**
- **By applying**
  - **Logic**
  - **Likely Outcome**
  - **Quantitative decision**





Decision Tree Structure

4



Node Splitting

5

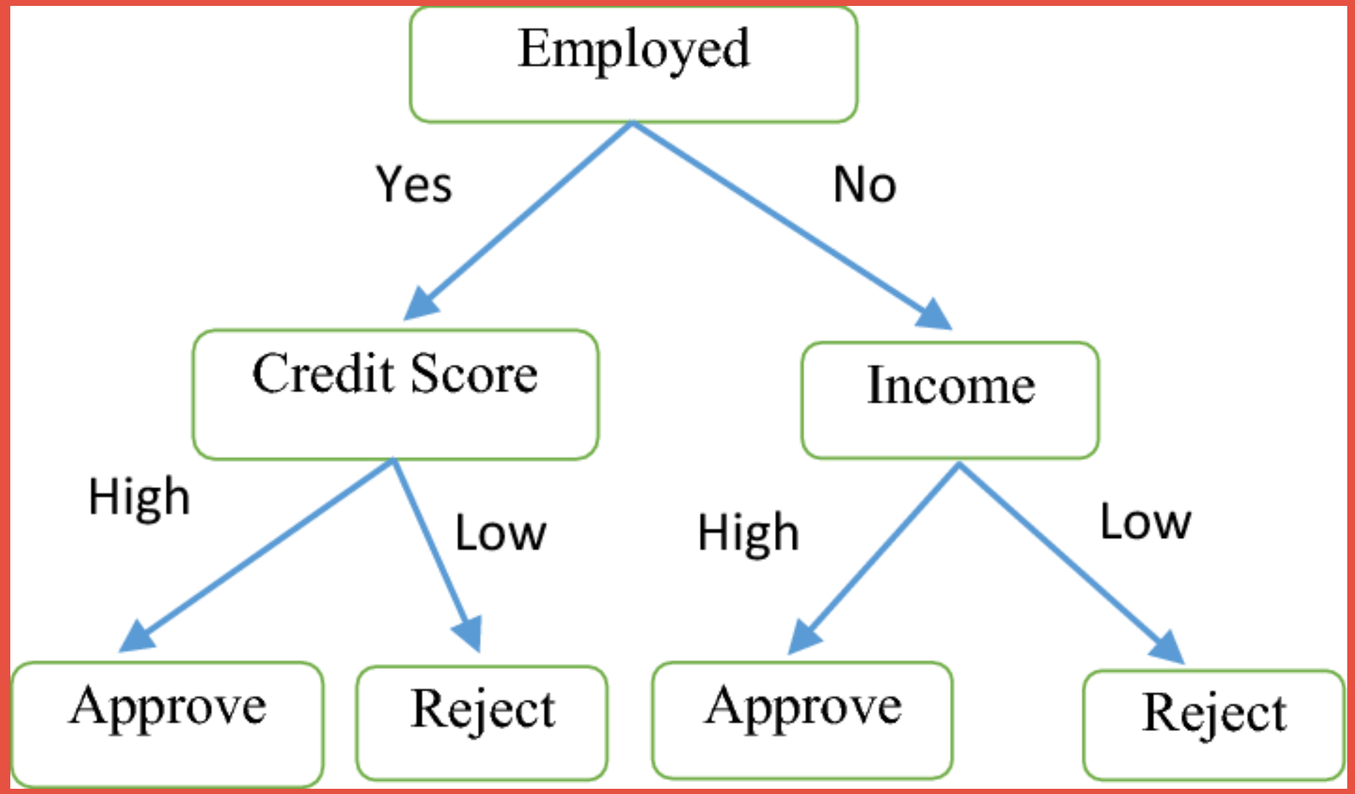
Importance of Node Splitting

- Effective dataset division
- Improved model performance
- Interpretability
- Segments dataset into homogeneous groups
- Captures complex decision boundaries
- Reduces bias and variance
- Enhances model interpretability

06

Example of Decision Tree

07

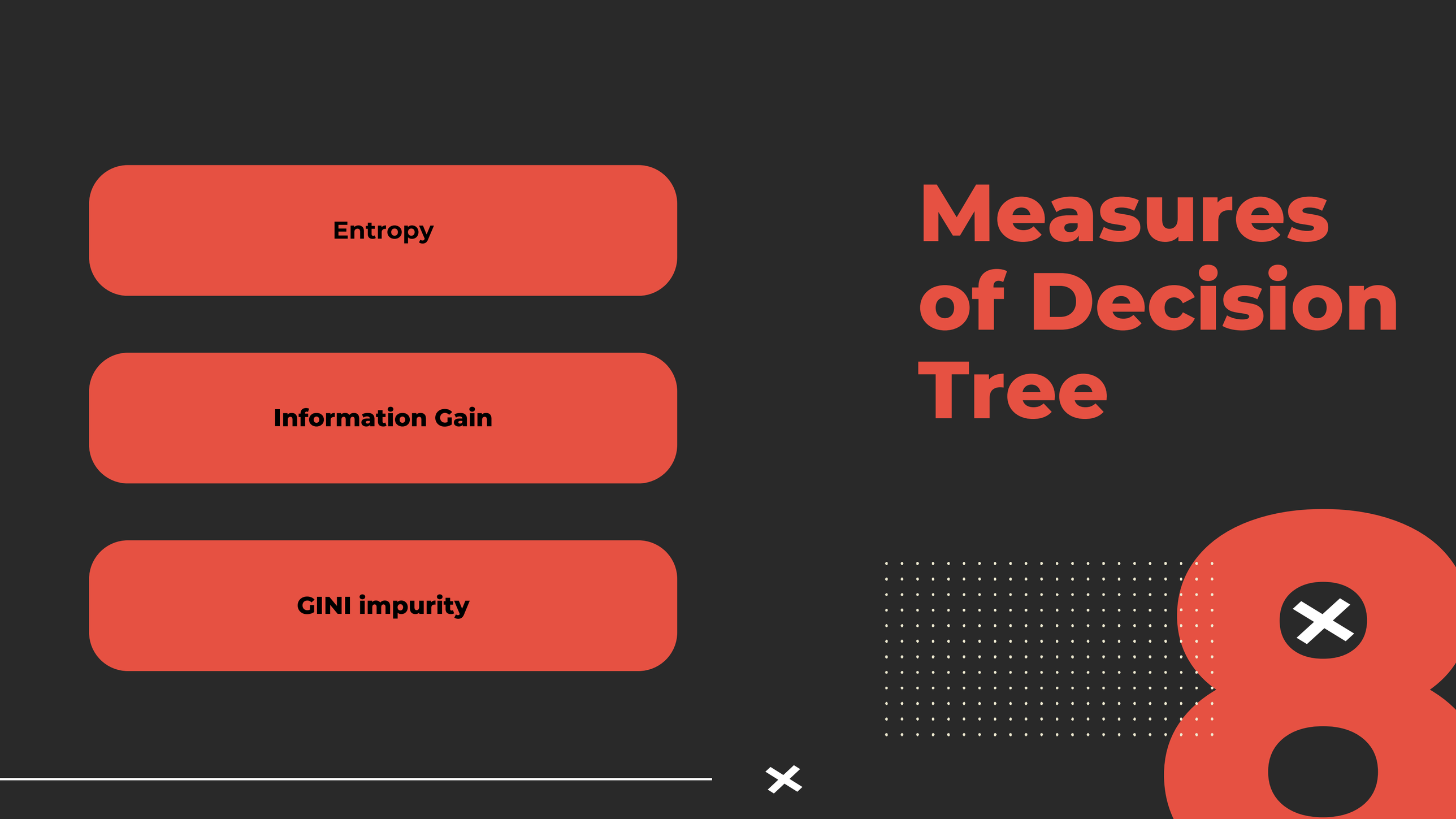


Measures of Decision Tree

Entropy

Information Gain

GINI impurity



What is Entropy?

Low Entropy

High Entropy

$$\text{Entropy} = \sum_{i=1}^{c} - p_i \log_2 p_i$$

c --> number of classes

p_i --> Probability of i^{th} class

Information Gain

$$\text{Information gain (T, F)} = \text{Entropy (T)} - \sum_{v \in F} \frac{|T_v|}{T} . \textit{Entropy}\,(T)$$

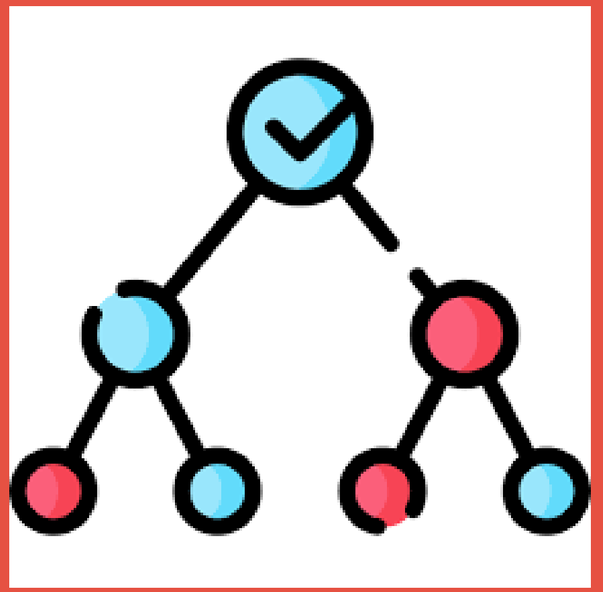
- Where T is target, F is the feature we're considering to split on, Tv is the subset of T for which Feature F has value v.

10

GINI Impurity

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

Where pi is the probability of an item being classified to a particular class.



TUTORIAL

12

Conclusion

- **Entropy** measures how mixed or uncertain our dataset's target variable is.
- **Information Gain** tells us how much a particular attribute helps in reducing the uncertainty about the target variable.
- **Gini Impurity** is another way to measure uncertainty in a dataset. It's similar to entropy but has a slightly different calculation.

- http://www.learnbymarketing.com/481/decision-tree-flavors-gini-info-gain/
- https://medium.com/analytics-steps/understanding-the-gini-index-and-information-gain-in-decision-trees-ab4720518ba8
- https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html
- https://towardsdatascience.com/understanding-decision-tree-classification-with-scikit-learn-2ddf272731bd
- https://www.shiksha.com/online-courses/articles/splitting-in-decision-tree/#:~:text=Splitting%20a%20single%20node%20into,concept%20that%20is%20diametrically%20opposite.
- https://gatesboltonanalytics.com/?page_id=282

References

THANK YOU!