

BrainStation Capstone Sprint 1

Ethan Alwaise

March 8, 2024

The Stack Exchange

A network of Q&A websites for various fields

The screenshot shows a Stack Overflow question page. The title "How exactly do decision trees split the input region?" is highlighted with a red border. The question was asked 24 days ago, modified 24 days ago, and viewed 176 times. The question text is: "Let's assume I have a rectangular input region with 3 points belonging to class 0 on the left and 1 point belonging to class 1 on the right. Let's assume these points are near the ends of the rectangle. In this scenario, how will the decision tree make the split? Will it split it right down the middle to offer 50% region to both sides, or will class 0 get 75% (3/4) of the region, leaving 25% (1/4) to class 1? Thanks in advance!". The question has 2 answers, with the top one tagged "decision-trees". The page includes a sidebar with navigation links (Home, Questions, Tags, Users, Companies, Unanswered, TEAMS), a "Stack Overflow for Teams" section, and a "Create a free Team" button. There are also advertisements by Google and a "The Overflow Blog" section on the right.

StackExchange Search on Data Science... Log in Sign up

How exactly do decision trees split the input region?

Asked 24 days ago Modified 24 days ago Viewed 176 times

Ads by Google

Send feedback Why this ad?

2

This is more of a stupid question!

Let's assume I have a rectangular input region with 3 points belonging to class 0 on the left and 1 point belonging to class 1 on the right. Let's assume these points are near the ends of the rectangle.

In this scenario, how will the decision tree make the split? Will it split it right down the middle to offer 50% region to both sides, or will class 0 get 75% (3/4) of the region, leaving 25% (1/4) to class 1?

Thanks in advance!

decision-trees

Share Improve this question Follow

asked Feb 12 at 13:03 Stella

The Overflow Blog

- Building GenAI features in practice with Intuit Mailchimp
- A leading ML educator on what you need to know about LLMs

Featured on Meta

- Our partnership with Google and commitment to socially responsible AI
- Shifting the data dump schedule: A proposal
- On AI-generated answers

Ads by Google

Send feedback

Why this ad?

The Problem

① Questions:

- What factors affect the likelihood that a question is answered?
- What patterns can we observe in posting activity over time?

② Who should care?

- Users who want to optimize their questions.
- Stakeholders in the Stack Exchange.

The Problem

- What factors affect the likelihood that a question is answered?
- What patterns can we observe in posting activity over time?

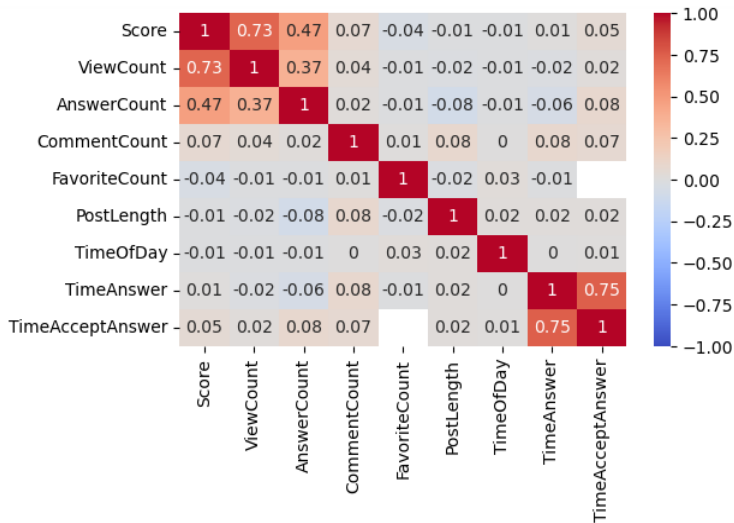
Approaches

- Binary classification (was the question answered?)
- Regression (how many answers did the question receive?)
- Time series analysis (can we forecast daily post numbers?)

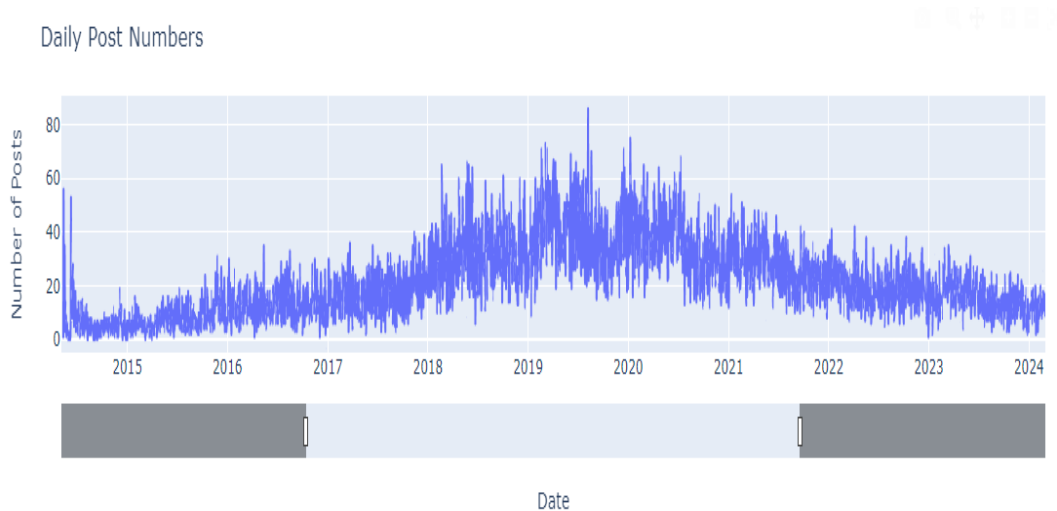
The Data

	PostType	CreationDate	Score	ViewCount	Body	LastActivityDate	Title	Tags	AnswerCount	CommentCount	LastEditDate
0	1	2014-05-13T23:58:30.457	9	959.0	<p>I've always been interested in machine lear...	2014-05-14T00:36:31.077	How can I do simple machine learning without h...	<machine-learning>	1.0	1	None
1	1	2014-05-14T00:11:06.457	4	503.0	<p>As a researcher and instructor, I'm looking...	2014-05-16T13:45:00.237	What open-source books (or other materials) pr...	<education> <open-source>	3.0	4	2014-05-16T13:45:00.237
2	2	2014-05-14T00:36:31.077	5	NaN	<p>Not sure if this fits the scope of this SE,...	2014-05-14T00:36:31.077	None	None	NaN	0	None
3	2	2014-05-14T00:53:43.273	13	NaN	<p>One book that's freely available is "The El...	2014-05-14T00:53:43.273	None	None	NaN	1	None
4	1	2014-05-14T01:25:59.677	26	1925.0	<p>I am sure data science as will be discussed...	2020-08-16T13:01:33.543	Is Data Science the Same as Data Mining?	<data-mining> <definitions>	4.0	1	2014-06-17T16:17:20.473

Correlations



Activity Over Time



Next Steps

- Process the text.
- Apply SARIMA to daily post numbers.