

Post Task Reflection

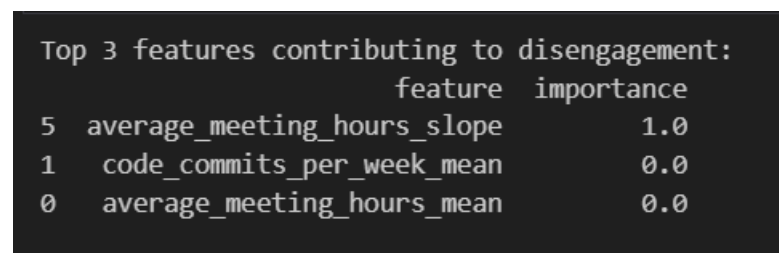
Training Model Used: Decision Tree

1. Top feature contributing to disengagement prediction:

I aggregated the **6-rows per employee** to a **single row per employee** by calculating the **mean and slope of each input feature**. Resulting in 2 features corresponding to a single weekly numeric feature.

*average_meeting_hours (6 rows) -> average_meeting_hours_mean ,
average_meeting_hours_slope (1 row)*

This resulted in a feature **average_meeting_hours_slope** which could easily classify all the given data and was the only feature on which decision tree was classifying the data.



```
Top 3 features contributing to disengagement:
      feature  importance
5  average_meeting_hours_slope      1.0
1  code_commits_per_week_mean       0.0
0  average_meeting_hours_mean       0.0
```

NOTE : This was due to the fact that the given data could be classified only on the basis of decreasing slope of the **average_meeting_hours** feature. If there had been variations in the data, then the decision tree would have needed more input features for the classification.

2. How did I process the time-series nature of data:

I aggregated the **6-rows per employee** to a **single row per employee** by calculating the **mean and slope of each input feature**. Resulting in 2 features corresponding to a single weekly numeric feature.

*average_meeting_hours (6 rows) -> average_meeting_hours_mean ,
average_meeting_hours_slope (1 row)*

The slope feature was needed because we had to classify on the basis of employee **gradually disengaging** and **slope was the best representation** of that. Now the decision tree could easily classify the data on the basis of these slopes (A highly negative slope indicating that the employee is disengaging from a particular activity).

3. What patterns did I notice among disengaging employees:

The disengaging employees had decreasing **average_meeting_hours**, **code_commits_per_week etc.** which indicated that they were slowly disengaging from work.

Average Meeting Hours Slope for a disengaging employee : -1.128571 (2nd employee)

Average Meeting Hours Slope for an engaged employee : 0.331429 (4th employee)

4. How would I ethically use this model:

I would:

- **Inform employees** that behavioral data (e.g. meeting hours, code commits, sentiment feedback) may be aggregated for insights into team health.
- **Obtain explicit consent** for this level of monitoring, and clearly document what data is collected and why

5. Improvement If I had access to more data:

I could have created more input features including **std. deviation, max, min values** of the input features for each employee while aggregating into a single row, but it would have **increased the dimensionality** a great extent and we didn't have that much data to train on. If we had more data with different variations, then including these extra features would have improved the performance further.