

# CPSC 330 Lecture 13: Feature importances

# Announcements

- HW4 grades are released
- HW5 is due next week Monday. Make use of office hours and tutorials this week.

# Scenario 1: Which model would you pick

Predicting whether a patient is likely to develop diabetes based on features such as age, blood pressure, glucose levels, and BMI. You have two models:

- LGBM which results in 0.9 f1 score
- Logistic regression which results in 0.84 f1 score

Which model would you pick? Why?

# Scenario 2

Predicting whether a user will purchase a product next based on their browsing history, previous purchases, and click behavior. You have two models:

- LGBM which results in 0.9 F1 score
- Logistic regression which results in 0.84 F1 score

Which model would you pick? Why?

# Transparency

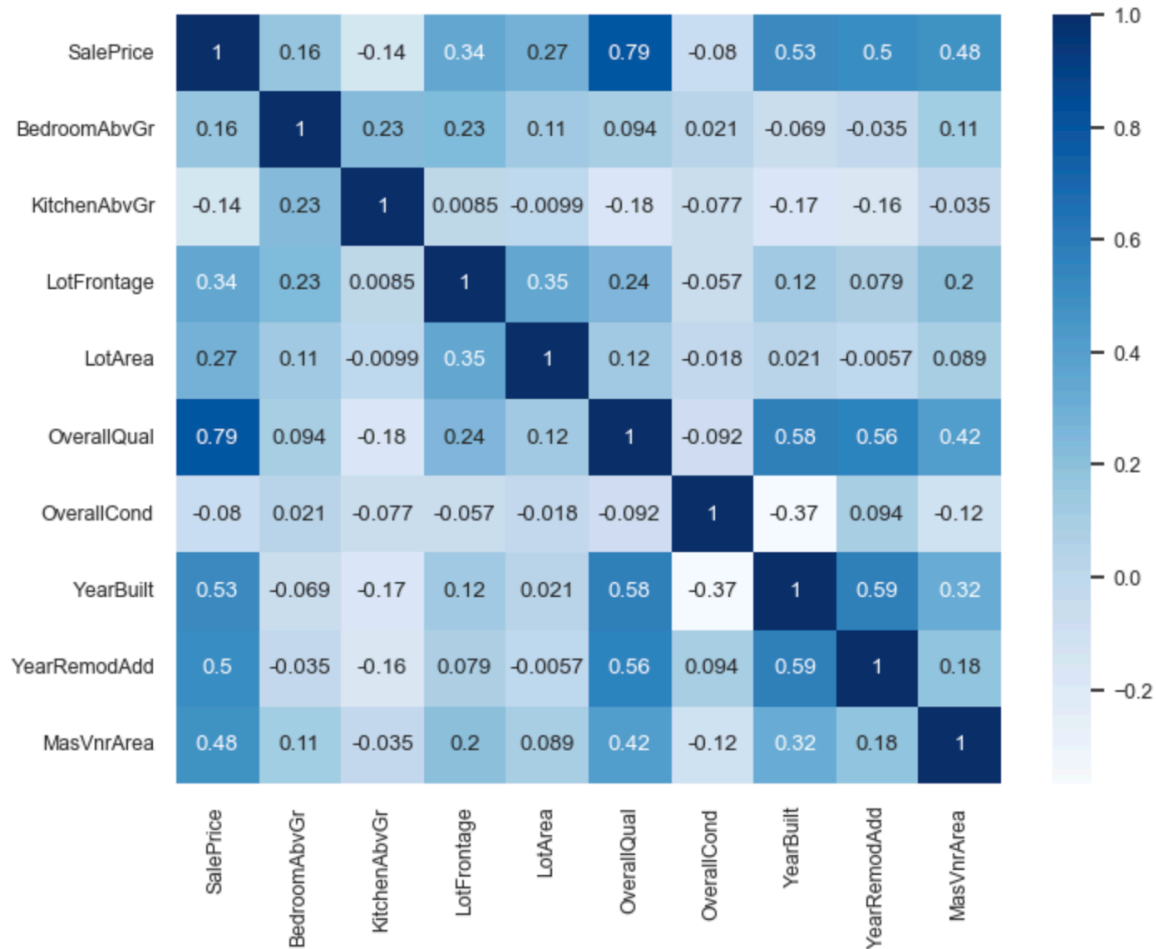
- In many domains understanding the relationship between features and predictions is critical for trust and regulatory compliance.

## Feature importances

- How does the output depend upon the input?
- How do the predictions change as a function of a particular feature?

# How to get feature importances?

# Correlations



- What are some limitations of correlations?

# Interpreting coefficients



# Interpreting coefficients

- When we have different types of preprocessed features, what challenges you might face in interpreting them?
  - Ordinally encoded features
  - One-hot encoded features
  - Scaled numeric features

# Group Work: Class Demo & Live Coding (if time permits)

For this demo, each student should [click this link](#) to create a new repo in their accounts, then clone that repo locally to follow along with the demo from today.

If you really don't want to create a repo,

- Navigate to the [cpsc330–2024W1](#) repo
- run `git pull` to pull the latest files in the course repo
- Look for the demo file [here](#)