

## **Data Science**

### **Inlämningsuppgift 3**

#### **Analys rapport grupparbete**

#### **Bakgrund**

Syftet med denna analys är att förutsäga hur många bostäder, flerbostadshus och småhus som förväntas byggas år 2029. Analysen ska undersöka vilka olika faktorer såsom län, flerbostadshus och småhus kommer att påverka bostadsbyggande år 2029. Denna prognos baseras på data från Bostadsenkät 2024.

#### **Genomförande**

##### **ETL process (Extract, Transform, Load)**

Första steget var att hämta data från en dataportal. I andra steget togs bort tomma rader och kolumner som var inte relevanta för analysen. Kolumnen Län var kategorisk och behövdes omvandla till numerisk värde. One- hot encoding användes för att skapa binära variabler för varje unikt län. Sista steget var att spara den transformerade data i Excel.

#### **Maskininlärning och Regression**

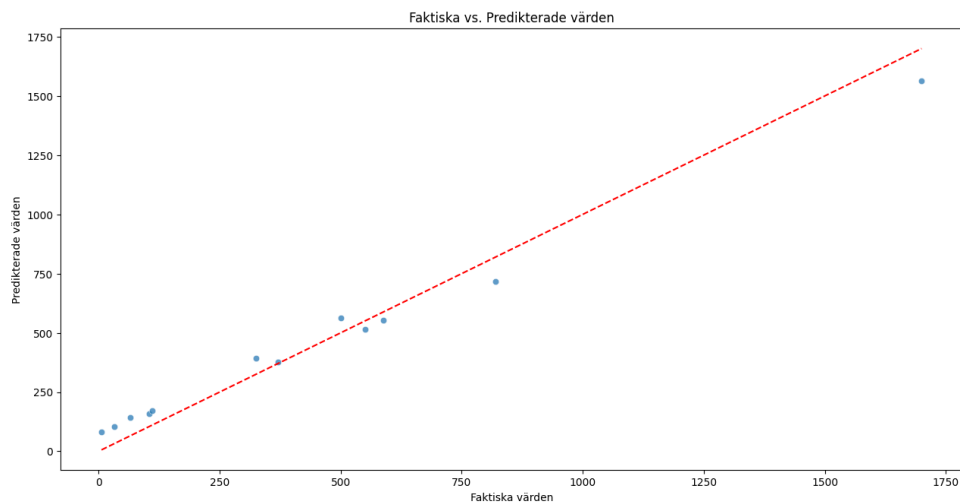
CatBoostRegressors regressionsmodell används för att förutsäga hur många bostäder som kommer att byggas år 2029. Modellen byggdes och tränades med data från 2024 till 2025. Sedan testades för att säkerställa att resultatet är så pålitligt som möjligt. Metoder som Mean absolute Error (MAE) och R2 används för att bedöma modellens prestanda. R2 score bedömer variation. X representerar oberoende variabler såsom län, flerbostadshus 2024, flerbostadshus 2025 , småbostadshus 2024, småbostadshus 2025. Målvariabel Y representerar totalt bostäder 2029.

#### **Resultat**

R2-score ligger på 0,97 vilket innebär att modellen är bra på att förutse och förklara hur många bostäder som kommer att byggas baserade på de oberoende variablerna. MAE ligger på 65,30, detta innebär att modellen gör 65 enheter fel varje gång den gör en förutsägelse.

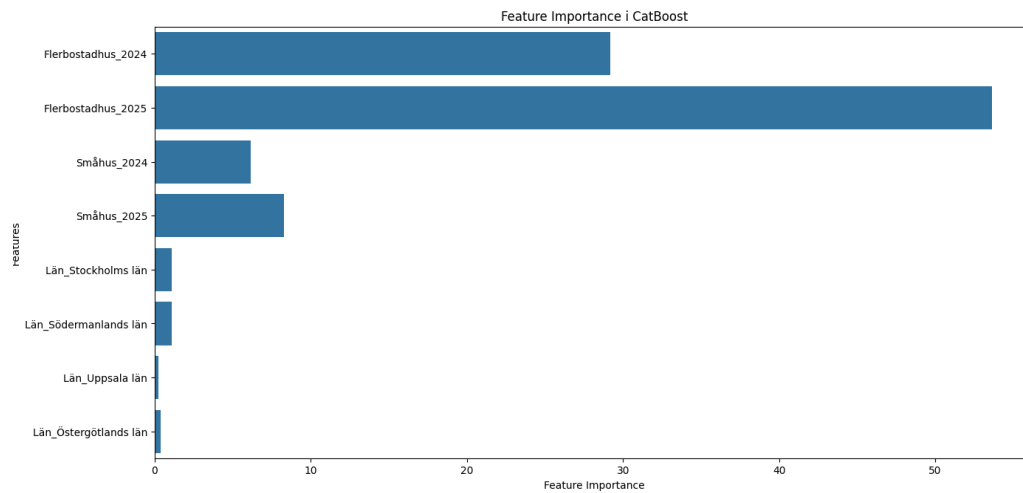
## Visualisering av resultat

### Diagram Scatter Plot



Scatterplot visar samband mellan de faktiska och predikterade värdena. Varje blå prick i diagram visar enskild jämförelse mellan det verkliga värde och modellens förutsägelse värde. Om alla punkterna skulle ligga exakt på prediktiva linje så skulle modellen vara perfekt och fungera utan fel men det finns vissa prickar som ligger under och ovanför linje. Prickar ovanför linje tyder på mer bostadsbyggande än förväntat 2029 medan prickar under linje visar mindre bostadsbyggande än förväntat.

### Diagram Barplot



Feature importance visas via barplot diagram för att möjliggöra att förstå hur mycket varje funktion bidrar till att bidra till modellens förmåga att göra förutsägelser. Flerbostadshus 2025 och flerbostadshus 2024 är de viktiga faktorer som har stor påverkan på prediktion. Småhus bostad 2024, småhus bostad 2025 har mindre betydelse än flerbostadshus 2024 och 2025.