

Förutspå bostadspriser

Namn: Husein Vilic
Projekt del 1

Problemformulering

Målet med projektet är att utveckla en maskininlärningsmodell som kan förutsäga bostadspriser baserat på olika faktorer. En bra modell skulle kunna hjälpa till att göra mer informerade beslut på bostadsmarknaden.

Mål i projektet

- Förutsäga genomsnittligt pris på bostäder i olika områden.
- Analysera hur olika faktorer (t.ex. inkomst, närhet till myndigheter, politisk ställning) påverkar priserna.
- Förstå hur dessa faktorer påverkar bostadspriser och ge klargöra vilka faktorer som är viktigast.

Datakällor

- Tidigare försäljningspriser, exempelvis från fastighetssajter, offentliga dataregister, så som mäklarstatistik, Boverket.
- Inkomst, kriminalitet, kollektivtrafik. Hämtas från exempelvis Statistiska Centralbyrån (SCB) samt andra nationella databaser, trafikverket.
- OpenStreetMap för geografiska bitarna samt lantmäteriet.
- Hemnet och Booli för aktuella priser.
- Folkhälsomyndigheten: Data om hälsorelaterade faktorer

Fält

- Bostadsdata: Priser, boyta, antal rum.
- Socioekonomiska faktorer: inkomst, utbildningsnivå, arbetslöshetsnivå, politisk ställning.
- Geografiska faktorer: tillgång till kollektivtrafik, kriminalitetsnivå, parker, andra rekreatiomsområden,

Datan

Datan kommer från pålitliga källor, Mäklarstatistik och SCB, vilka båda är pålitliga informationskällor för bostads- och socioekonomisk data. Datan är omfattande och täcker relevanta faktorer, vilket ger en god grund för att bygga en noggrann och robust modell för bostadsprisförutsägelse.

Datan kommer från pålitliga källor, vilket säkerställer hög kvalitet och relevans. Slutpriser hämtas från fastighetssajter och Mäklarstatistik. Detta ger en stabil grund för historisk prisutveckling. Socioekonomisk information, som inkomst, kriminalitet politisk ställning, kommer från Statistiska Centralbyrån (SCB).

Geografiska data samlas in från OpenStreetMap och Lantmäteriet, som ger information om viktiga faciliteter. Aktuella priser kan hämtas från Hemnet och Booli, medan Folkhälsomyndigheten bidrar med data om hälsorelaterade faktorer.

Modeller

Eventuella modeller att överväga:

- Linjär regression
- Gradient Boosting
- Beslutsstödträd och Random Forest
- Neurala nätverk

Projektet faller inom regressionsproblem

- Förutsäga ett kontinuerligt värde. Bostadspriser kan variera, vilket faller inom regressionsproblem.
- Oberoende variabler: Finns olika faktorer (t.ex. tidigare försäljningspriser, socioekonomiska data, geografiska data, hälsorelaterade faktorer)

Resultatet vi vill åstadkomma

Modellen ska kunna ta socioekonomiska och geografiska data, samt bostadsdata för ett område och förutsäga bostadspriser med så hög noggrannhet som möjligt. Vi ska också kunna analysera vilka faktorer som påverkar bostadspriserna mest, och presentera dessa resultat för bättre insikter om bostadsmarknaden.