

The background features a light pink base with various decorative elements. In the top left, there is a blue line-art floral motif. In the top right, a large blue abstract shape with white wavy lines and orange dots is visible. In the bottom left, an orange abstract shape with a yellow rounded rectangle and orange dots is present. In the bottom right, a pink line-art floral motif is shown. A dashed line arches over the top of the title text.

ANALYS OCH PREDIKTION AV BOSTADSPRISER MED MASKININLÄRNING

Alia Atawna
Examensarbete



AGENDA

- Inledning
- Syfte och frågeställningar
- Metod och dataset
- Maskininlärningsmodeller
- Resultat och Diskussion
- Begränsningar och framtida forskning
- Praktiska tillämpningar
- Slutsatser

INLEDNING

- Maskininlärning har revolutionerat många områden, inklusive fastighetsmarknaden.
- Målet är att använda dataanalys för att förutsäga fastighetspriser och förstå marknadsdynamik.



SYFTE OCH FRÅGESTÄLLNINGAR

Att undersöka hur maskininlärning kan förutsäga fastighetspriser och identifiera de mest betydelsefulla faktorerna för värderingen.

1. Vilka faktorer påverkar fastighetspriser mest?
2. Hur presterar olika maskininlärningsmodeller?
3. Vilka insikter kan erbjudas fastighetsägare och investerare?

METOD OCH DATASET

- Kodning av variabler
- Log-transformering
- Hantering av outliers

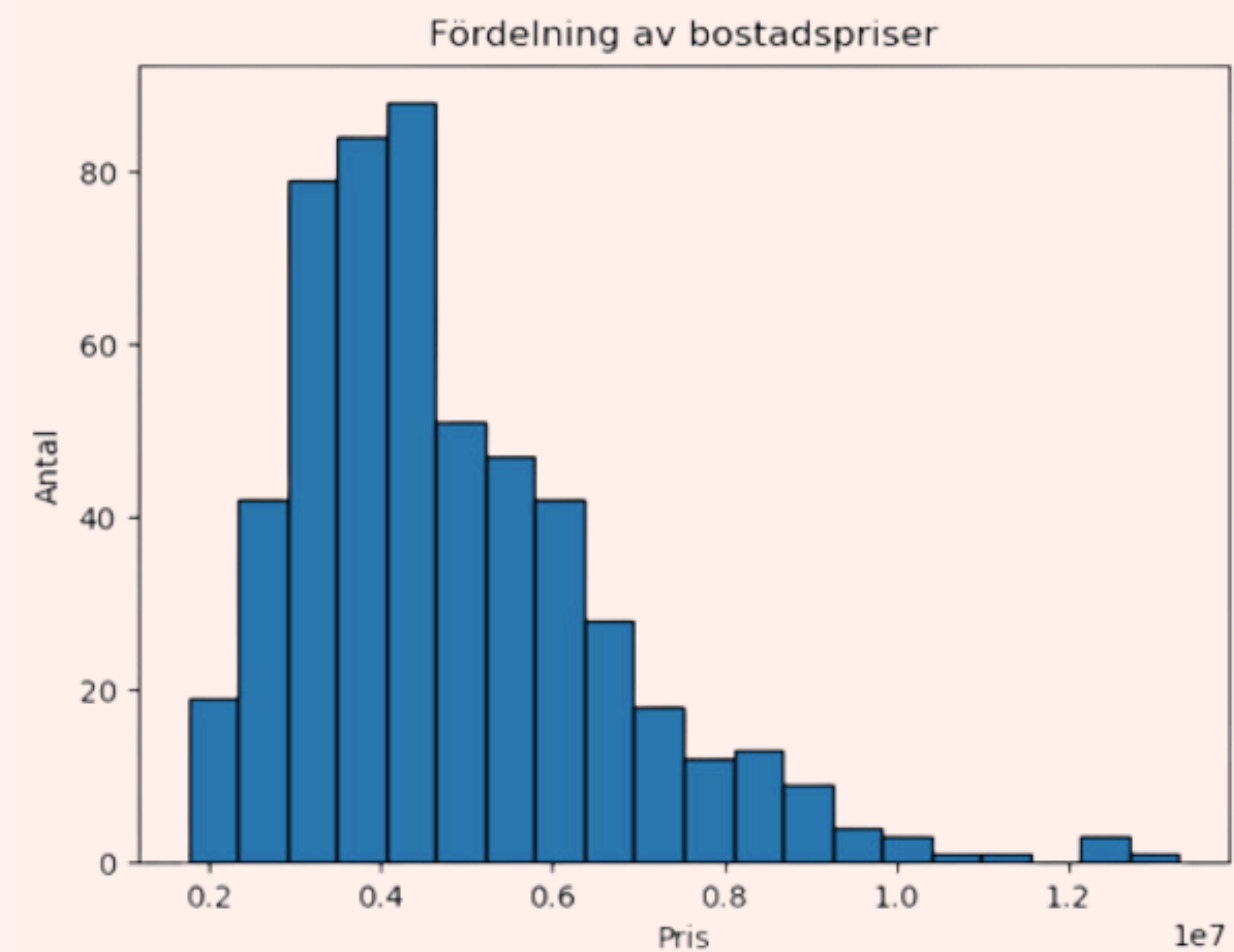
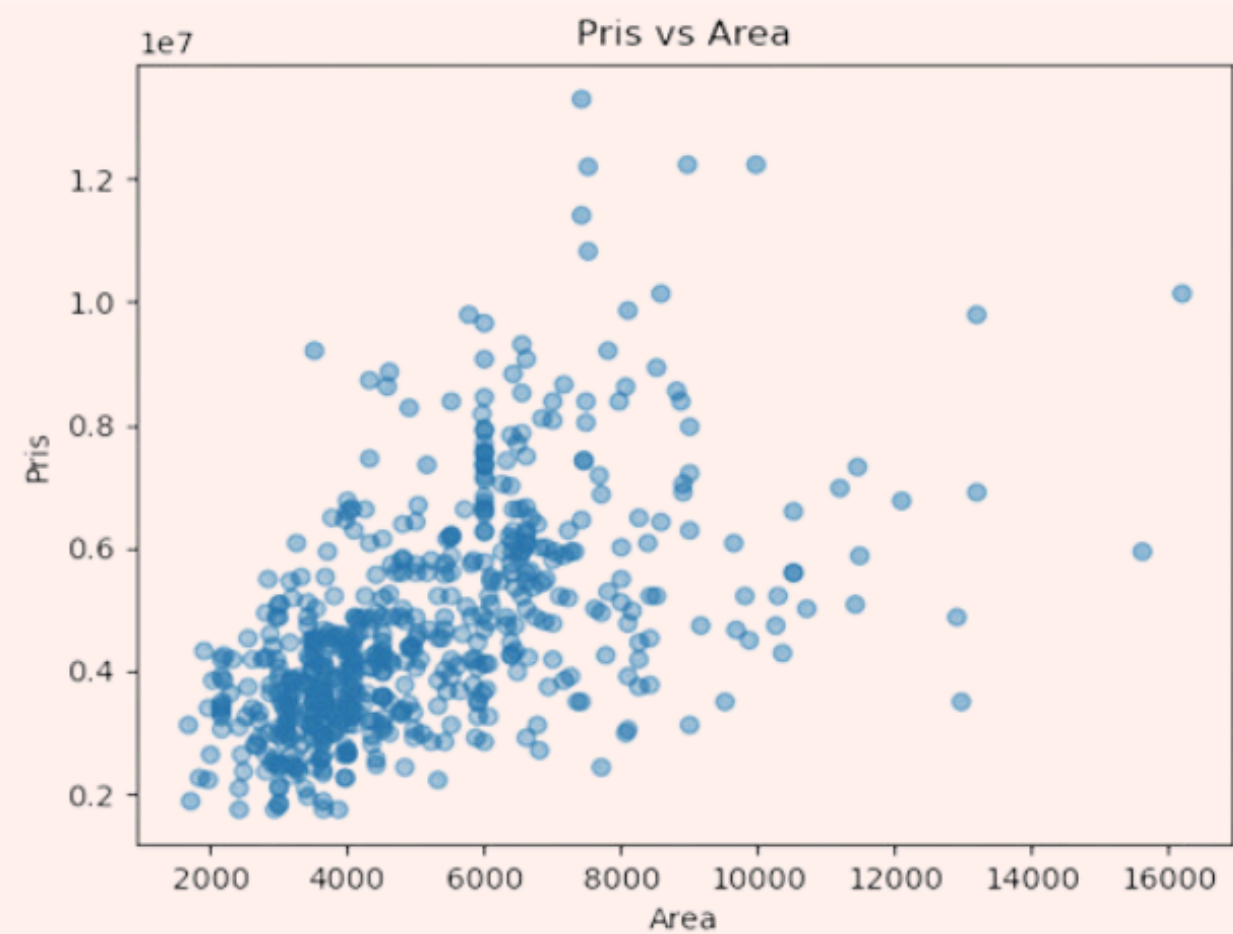
Dataset med 545 observationer:
Pris, area, sovrum, badrum, med mera.

De första raderna av datasetet:

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement
0	13300000	7420	4	2	3	yes	no	no
1	12250000	8960	4	4	4	yes	no	no
2	12250000	9960	3	2	2	yes	no	yes
3	12215000	7500	4	2	2	yes	no	yes
4	11410000	7420	4	1	2	yes	yes	yes

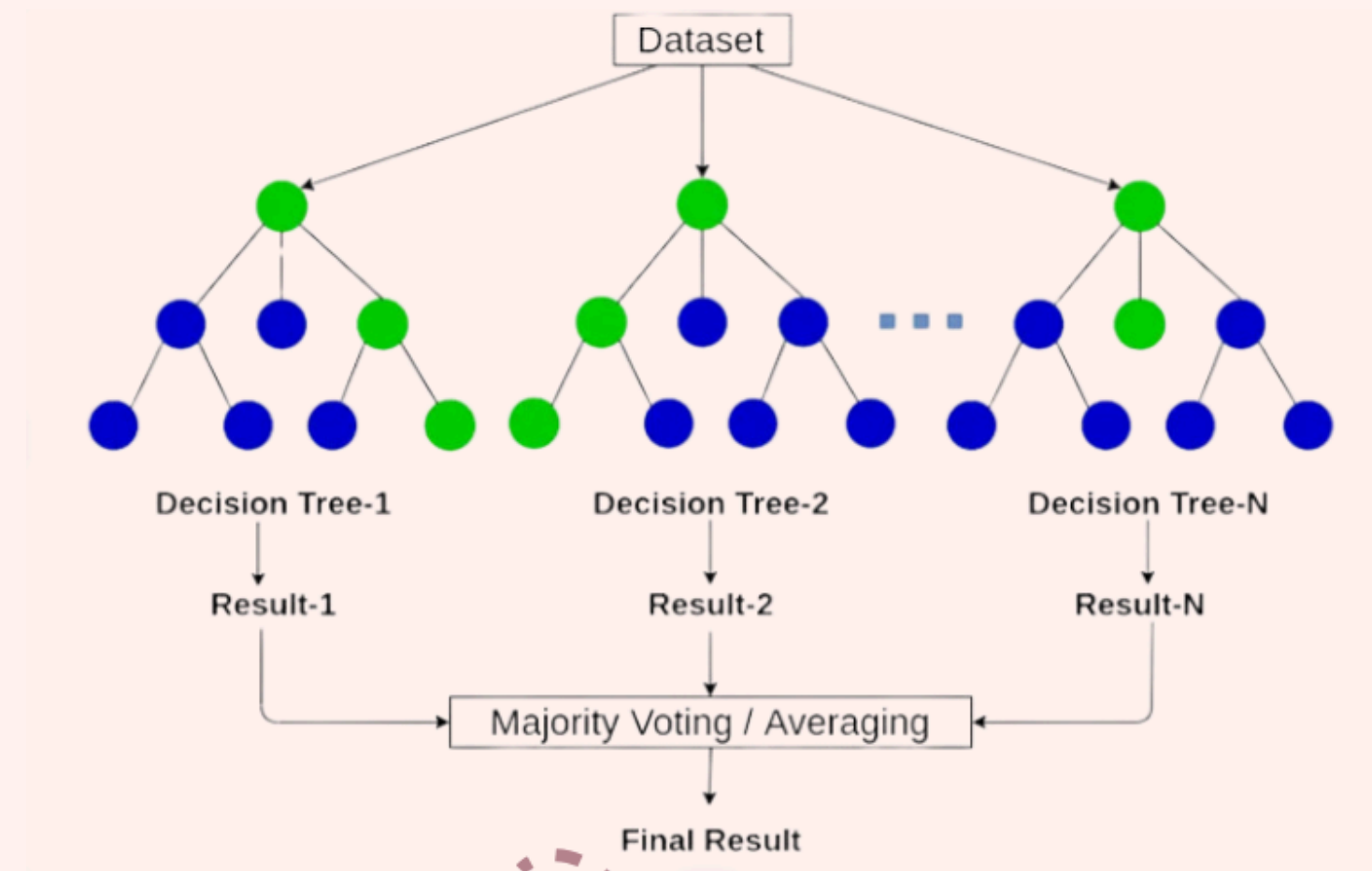
	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
0	no	yes	2	yes	furnished
1	no	yes	3	no	furnished
2	no	no	2	yes	semi-furnished
3	no	yes	3	yes	furnished
4	no	yes	2	no	furnished

UTFORSKANDE DATAANALYS (EDA)



IMPLEMENTERADE MODELLER

- **Linjär regression:** För linjära samband.
- **Random Forest:** För att identifiera viktiga variabler.
- **SVR:** För komplexa, icke-linjära samband.



MODELLERNAS PRESTANDA

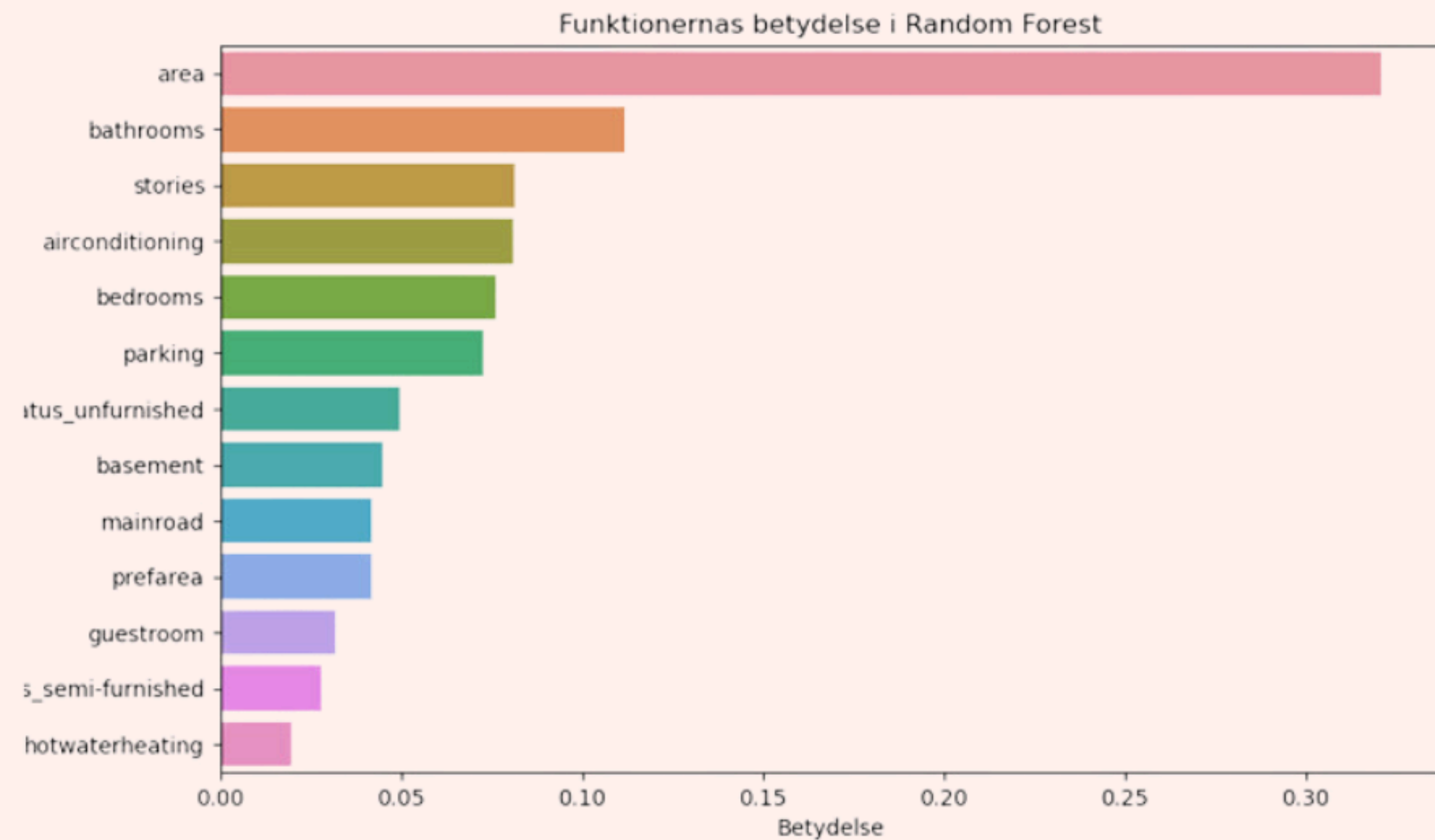
SVR presterade bäst

MODELL	RMSE	MAE	R ²
ENKEL LINJÄR REGRESSION	1,039,104.19	791,076.37	0.72
RANDOM FOREST	1,192,203.00	851,407.32	0.63
SVR	230,000.00	180,000.00	0.68

VARIABLERNAS BETYDELSE (RANDOM FOREST)

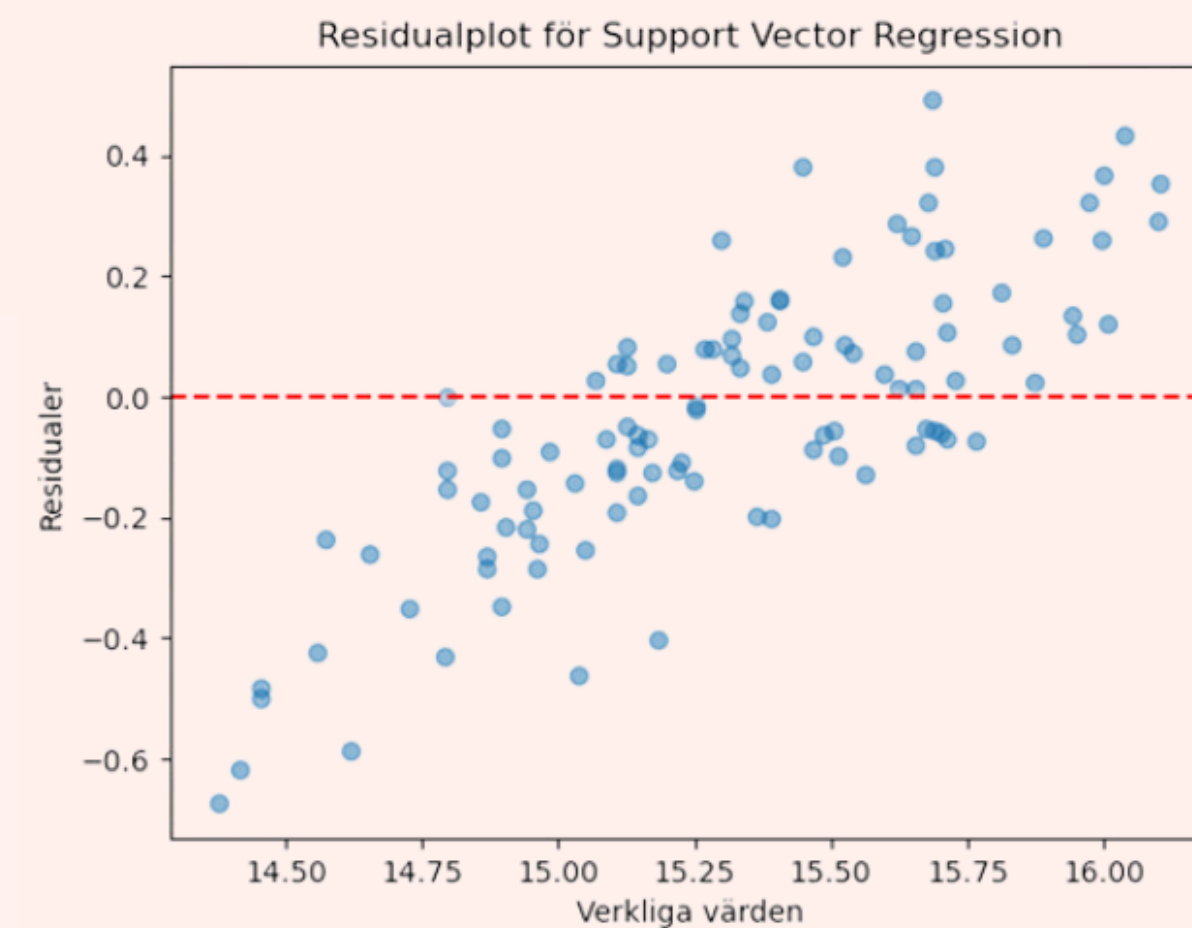
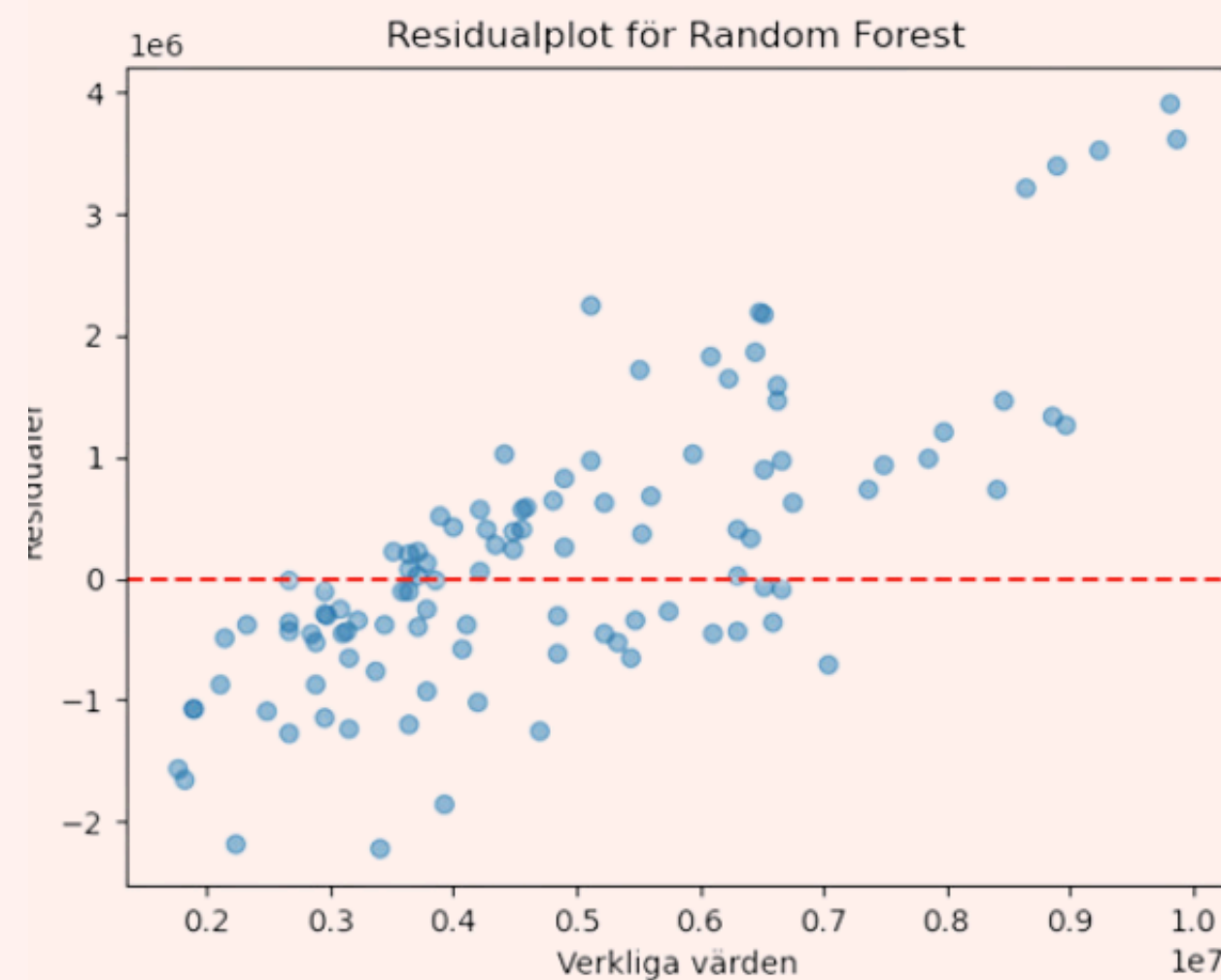
Viktigaste variablerna:

- Bostadsarea
- Badrum
- Våningsplan



DISKUSSION

- Styrkor och begränsningar av varje modell.
- Etiska aspekter: Integritet, bias.



BEGRÄNSNINGAR OCH FRAMTIDA FORSKNING

Begränsningar:

- Litet dataset
- Ingen geografisk variation

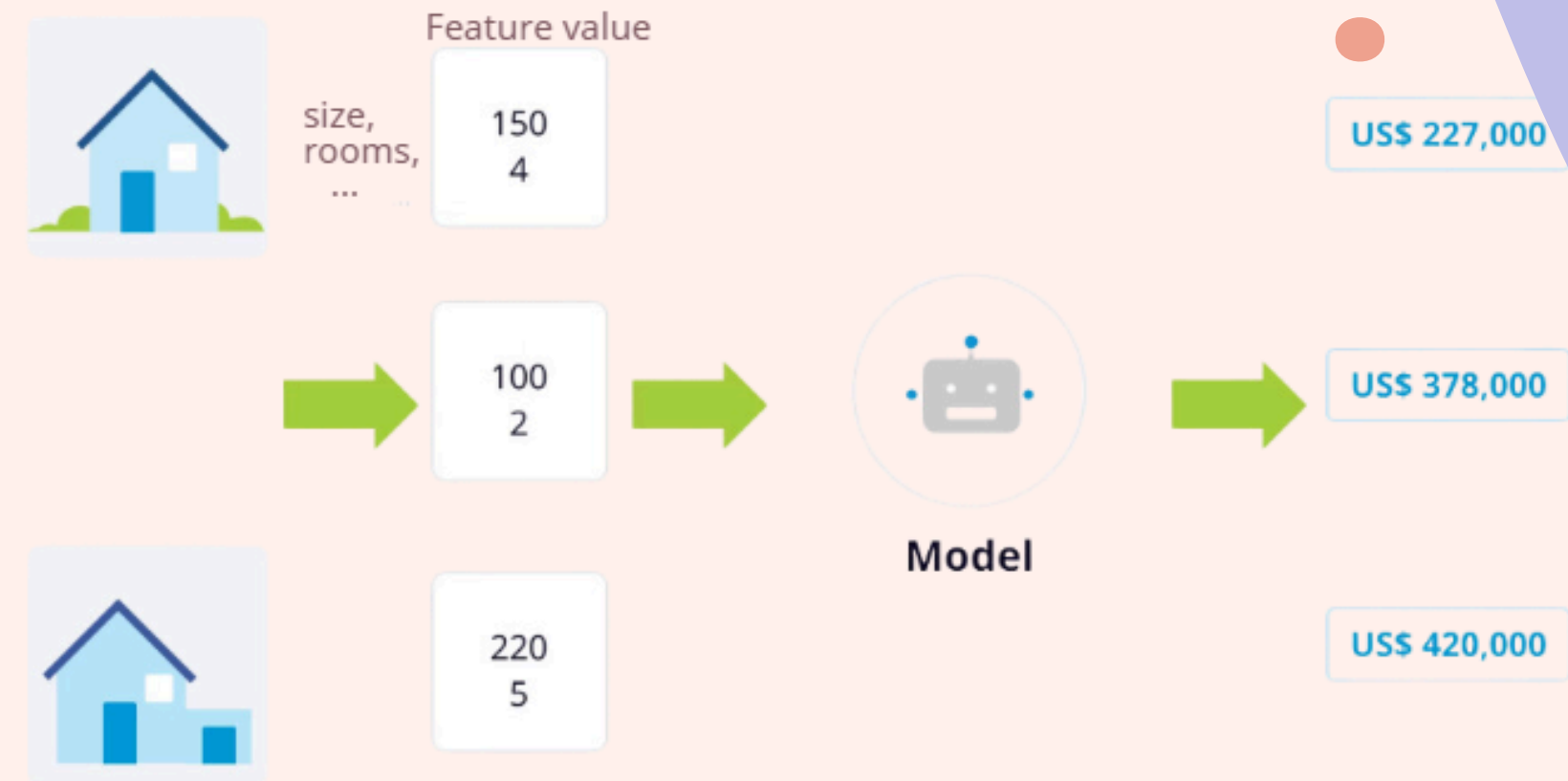
Framtid:

- Integrera text
- Bilder
- Realtidsdata



PRAKTISKA TILLÄMPNINGAR

- Insikter för fastighetsägare, investerare, stadsplanerare.
- Automatiserade värderingssystem.



SLUTSATSER

- SVR bäst för icke-linjära samband.
- Random Forest ger viktiga insikter om variabler.
- Potential för framtida forskning och tillämpningar.





TACK!