Prognostisering av framtida utbildningskostnader



Lidiia Kashevarova

EC Utbildning

Kostnader för grundskola och

gymnasieskola under år 2025-2035

202410

# Abstract

Abstract

The objective of this study is to forecast the costs of education in middle and high schools across Swedish regions for the period 2025 to 2035. The forecast is based on both actual birth data from 2007 to 2023, sourced from Statistics Sweden (SCB), and projected values for 2024 to 2029, calculated using an LSTM model built with Keras. A validation of the projected birth data against real values from SCB statistics yielded the following average percentage errors for each year:

2019: 2.32%

2020: 4.04%

2021: 3.43%

2022: 8.90%

2023: 12.90%

The number of children expected to attend school in each analyzed year was estimated based on birth rates, mortality rates (calculated for each age group using data from "Döda efter region, ålder och år"), and migration statistics (derived from "Flyttningsöverskott efter region, ålder och år"). The mortality rates and migration trends were analyzed by age and region, ensuring a comprehensive and region-specific forecast.

To estimate the cost of education for middle and high school students over the next 10 years, data from the "Kostnader för grundskola och gymnasieskola efter skolform, huvudman och kostnadsslag" (2007–2022) dataset was used. This dataset provides the fixed and current costs per student. Linear regression was employed to forecast the education costs for the years 2025 to 2035.

Innehållsförteckning

[1 Inledning 1](#_Toc176961624)

[2 Teori 2](#_Toc176961626)

[2.1 LSTM-modell 3](#_Toc176961627)

[3 Metod 4](#_Toc176961632)

[3.1 Data 4](#_Toc176961633)

[3.2 Beräkning av dödlighet och migrationsprocent 3](#_Toc176961634)

3.3 SQL-databas………………………………………………………………………………………………………………………6

3.4 Prognostisering av antalet födda barn för åren 2024–2029 med hjälp av LSTM-modellen…6

3.5 Beräkning av antalet elever från år 2025 till 2035……………………………………………………………..8

3.6 Prognostisering av kostnaden för utbildning per elev med hjälp av Linjär Regression……….9

3.7 Beräkning av den totala utbildningskostnaden per region och år……………………………………10

[4 Resultat och Diskussion 11](#_Toc176961635)

4.1 Streamlit………………………………………………………………………………………………………………………….11

4.2 Diskussion……………………………………………………………………………………………………………………….14

[5 Självutvärdering 15](#_Toc176961637)

[Källförteckning 16](#_Toc176961639)

# Inledning

Prognostisering av framtida utbildningskostnader är en viktig fråga för att kunna allokera resurser på ett effektivt sätt, men det primära fokuset i detta arbete ligger på de metoder och verktyg som har använts för att genomföra prognoserna. Denna studie har gett mig möjlighet att tillämpa och jämföra olika prognostiseringsmetoder, såsom maskininlärningsmodeller (LSTM) och traditionella statistiska metoder, samt att hantera och bearbeta stora datamängder från flera källor.

Syfte och frågeställningar

Syftet med denna rapport är att tillämpa och utvärdera olika prognosmetoder för att förutsäga utbildningskostnader i grundskolan och gymnasieskolan i svenska regioner för perioden 2025–2035. För att uppfylla syftet kommer följande frågeställningar att besvaras:

* 1. Hur kan maskininlärningsmodeller användas för att prognostisera antalet barn i skolåldern för kommande år?
  2. Hur kan traditionella statistiska metoder bidra till att skapa prognoser för framtida utbildningskostnader per region?

# Teori

## LSTM-modell

”RNN och LSTM är speciella neurala nätverksarkitekturer som kan bearbeta sekventiell data, data där kronologisk ordning spelar roll dock LSTM (Long Short-Term Memory) har mer avancerat ”minne”. (Nelson, Internet)

”LSTM noder har både ett korttidsminne () och ett långtidsminne (). Långtidsminnet glömmer vissa minnen och lägger till andra.

* och används i fyra olika komponenter.
* analyserar input och .
* ”Gate Controllers” *, ,* använder logistisk aktiveringsfunktion villket innebär att outputen är mellan [0, 1] och används för elementvis-multiplikation. Om ”gate controller” = 0 🡪 ”de stänger porten”.

1. styr vilken del av långtidsminnet som skall glömmas.
2. styr vilken del av som skall läggas till långtidsminnet.
3. styr vilken del av som skall användas för och .

LSTM celler är flexibla i termer att de kan lära sig vad de skall ha kvar i långtidsminnet, vad som skall glömmas bort och vad den skall använda. LSTM har varit en framgångsrik modell i tillämpningsområden såsom tidsserier, längre texter, ljudinspelningar och sekventiell data i allmänhet.” (Prgomet, GitHub)

En bild som visar text, diagram, linje, Plan

Automatiskt genererad beskrivning

*Figur 1: LSTM (Long Short-Term Memory)(*  *Géron, A., s.516].*

” LSTM är väsentligt förbättrade versioner av RNN, kapabel att tolka längre sekvenser av data. LSTM-modeller består av tre olika komponenter, eller grindar. Det är en ingångsgrind, en utgångsgrind och en glömgrind. LSTM:er hänsyn till indata från föregående tidssteg när man ändrar modellens minne och ingångsvikter. Ingångsporten fattar beslut om vilka värden som är viktiga och bör släppas igenom modellen. En sigmoidfunktion används i ingångsgrinden, som gör bestämningar om vilka värden som ska skickas vidare genom det återkommande nätverket. Noll sänker värdet, medan 1 bevarar det. En TanH-funktion används också här, som avgör hur viktiga ingångsvärdena är för modellen, från -1 till 1.”(Nilson, Internet)

### 2.2. Linjär Regression

### ”Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable..” (IBM, Interenet)

**Formeln för linjär regression**

Formellt kan linjär regressionsmodellen representeras enligt följande:

”Y=β0+β1X1+β2X2+...+βnXn+ϵY

där:

* YYY — den beroende variabeln (utbildningskostnad),
* β0​ — konstanten,
* β1,β2,...,βn — koefficienter som visar hur mycket varje oberoende variabel påverkar den beroende,
* X1,X2,...,Xn ​ — oberoende variabler (faktorer som påverkar kostnaden),

ϵ — modellens fel.” (GeeksforGeeks , Internet)

Linjär regression bygger på antagandet att det finns ett linjärt samband mellan en oberoende variabel och en beroende variabel. Genom att analysera detta samband kan linjär regression användas för att förutsäga framtida värden baserat på tidigare observationer.

# Metod

### 3.1. Data

I detta projekt användes följande data som hämtades via API från SCB webbplats:

”1. [Levande födda efter region och år. År 1968 - 2023](https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101H/FoddaK/)

2. [Folkmängd 0-19 år efter region, ålder och år](https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101A/BefolkningNy/table/tableViewLayout1/)

3. [Döda efter region, ålder och år](https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101I/DodaFodelsearK/)

4. [Flyttningsöverskott efter region, ålder och år](https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101J/Flyttningar97/table/tableViewLayout1/)

5. [Kostnader för grundskola och gymnasieskola efter skolform, huvudman och kostnadsslag. År 2007 - 2022](https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__UF__UF0514/UtbKostGrundGym/) ”, (SCB, Internet)

I den här studien förutspås kostnaden för utbildning i grundskola och gymnasium under perioden 2025 till 2035. Barn börjar vanligtvis i grundskolan vid 6 års ålder och avslutar gymnasiet vid 18 års ålder. Därför fokuserar vi på barn födda mellan 2007 och 2029.

### 3.2. Beräkning av dödlighet och migrationsprocent

För att ta hänsyn till förändringar i befolkningsstorleken för varje ålder beräknades dödlighetsprocenten för varje åldersgrupp från 0 till 18 år (baserat på statistiken "Döda efter region, ålder och år") samt migrationsprocenten för varje ålder och region (baserat på statistiken "Flyttningsöverskott efter region, ålder och år") i förhållande till den totala populationen (data från statistiken "Folkmängd 0-19 år efter region, ålder och år").

|  |  |
| --- | --- |
| Table mortality\_rate:  Age Mortality\_Rate  0 0 0.212834  1 1 0.043071  2 2 0.015785  3 3 0.011063  4 4 0.009583  5 5 0.008885  6 6 0.007945  7 7 0.006961  8 8 0.006620  9 9 0.006604  10 10 0.006516  11 11 0.007322  12 12 0.007257  13 13 0.009609  14 14 0.010670  15 15 0.015036  16 16 0.016584  17 17 0.022521  18 18 0.027203  19 19 0.035517 | Table migration\_rate:  Region Age Average\_Migration\_Rate  0 1 0 0.029957  1 1 1 0.029345  2 1 2 -0.122223  3 1 3 -0.191543  4 1 4 -0.066815  5 1 5 -0.001160  6 1 6 0.029916  7 1 7 0.165604  8 1 8 0.171212  9 1 9 0.304531  10 1 10 0.350064  11 1 11 0.428321  12 1 12 0.483905  13 1 13 0.575142  14 1 14 0.659856  15 1 15 0.775383  16 1 16 1.186470  17 1 17 1.199185  18 1 18 1.404621  19 1 19 1.230151 |

*Tabell 1: Exempel på beräknade dödlighets- och migrationsprocenter av den totala populationen för respektive åldersgrupp.*

För att kontrollera riktigheten av de beräknade dödlighets- och migrationsprocenterna genomfördes följande kontroll. Vi använder följande ekvation:

Antal\_barn\_n = Antal\_födda\_n-1 \* (1 - mortality\_rate\_n / 100) \* (1 + migration\_rate\_n / 100) *(ekv.1)*

Där:

* Antal\_barn\_n är antalet barn vid ålder n.
* Antal\_födda\_n-1 är antalet barn födda vid föregående ålder (n−1).
* mortality\_rate\_n är dödlighetskoefficienten för ålder n.
* migration\_rate\_n är migrationskoefficienten för ålder n.

Som utgångspunkt togs antalet barn födda år 2004, och detta antal omräknades enligt *ekvation1* med hänsyn till dödlighets- och migrationskoefficienterna för åldern 0 år gammal. Därefter omräknades det beräknade värdet för åldern 0 enligt *ekvation1* för åldern 1 år gammal. Denna process upprepades fram till 19 års ålder, där det beräknade värdet för varje tidigare ålder omräknades enligt *ekvation1* för att beräkna antalet barn vid nästa ålder.

De beräknade värd jämfördes med de faktiska antalen från statistiken "Folkmängd 0-19 år efter region, ålder och år" för motsvarande år och region och den absoluta och relativa skillnaden mellan dessa värden räknades ut.

|  |  |  |  |
| --- | --- | --- | --- |
| Ålder | Stockholms län | Skåne län | Västra Götalands län |
| 0  1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19 | -0.36  -0.35  -0.32  -0.09  0.11  0.66  1.01  1.26  1.60  1.79  2.09  2.10  2.32  2.67  3.09  3.07  2.58  2.00  1.61  0.05 | -0.12  -0.22  0.54  1.05  1.59  1.87  1.53  1.35  0.90  0.96  0.99  1.67  2.64  3.65  4.48  4.59  4.24  3.79  3.64  3.25 | -0.39  -0.35  -0.60  -0.82  -1.03  -1.06  -1.07  -1.16  -1.03  -0.28  0.22  0.79  1.80  2.44  3.05  2.95  2.29  1.93  1.48  1.86 |

*Tabell 2: Exempel på relativ skillnad mellan det beräknade och det faktiska antalet barn i tre regioner, %*

Generellt sett är skillnaden inte betydande. Den största relativa skillnaden mellan det beräknade och det faktiska antalet personer var 9 % i Gotlands län för åldern 8 år, men vid 13 års ålder hade skillnaden minskat och översteg inte 4 % fram till 19 års ålder. Detta leder till slutsatsen att dödlighets- och migrationskoefficienterna möjliggör ganska tillförlitlig beräkning av antalet personer i varje ålder, baserat på antalet födda barn.

### 3.3. SQL-databas

För att underlätta lagring och användning av data skapades en SQL-databas, 'education\_costs\_db.sqlite'. Denna databas innehåller all ursprunglig data hämtad från SCB statistik samt de beräknade data som genererades under programmets körning.

### 3.4. Prognostisering av antalet födda barn för åren 2024–2029 med hjälp av LSTM-modellen

I detta projekt användes LSTM-modellen, som är en del av deep learning API Keras. Valet av LSTM baserades på det faktum att andra modeller, som Random Forest Regressor, ARIMA och SARIMA, trots hög noggrannhet i mätningarna (MSE, MAE, R²), visade sig vara okänsliga för förändringar i antalet födda barn med år inom en region. Dessa modeller fokuserade främst på förändringar i befolkningsantalet mellan regioner, vilket resulterade i att prognoserna för antalet födda barn mellan 2024 och 2029 förblev konstanta inom en region. Detta begränsade modellernas förmåga att anpassa sig till verkliga demografiska trender. För att skapa mer dynamiska och realistiska prognoser valdes därför LSTM-modellen, som hanterar sekventiella data och långsiktiga beroenden bättre.

LSTM-modellen användes för att förutsäga antalet barn som föds under perioden 2024 till 2029. Modellen tränades på data hämtade från SCB statistik "Levande födda efter region och år" för åren 1968 till 2018. Data för åren 2019-2023 användes för att testa modellens prestanda och bedöma dess prognosnoggrannhet.

|  |
| --- |
| Percentage error for each region and year: |
| Region 1 3 4 5 6 7 \ |
| 2019 16.467156 17.199467 11.975534 9.047391 9.813856 17.404508 |
| 2020 16.880452 14.673942 8.670428 5.818341 11.390861 19.012911 |
| 2021 18.363131 16.943772 9.699246 9.292191 8.663118 16.081707 |
| 2022 9.469425 11.261526 4.473130 0.581114 0.972666 6.891674 |
| 2023 7.443943 7.434082 8.004217 5.643879 2.454012 3.770542 |
|  |
| Region 8 9 10 12 ... 14 17 \ |
| 2019 3.660121 1.164529 2.355760 15.014966 ... 9.790531 5.399372 |
| 2020 7.307790 9.264601 0.998504 14.000009 ... 9.070754 3.147583 |
| 2021 6.045765 2.422499 0.214251 14.548735 ... 10.497396 0.041061 |
| 2022 5.603835 9.140041 7.433012 7.022246 ... 3.038693 3.912595 |
| 2023 7.463152 10.861334 14.185274 3.414166 ... 0.495460 14.713366 |
|  |
| Region 18 19 20 21 22 23 \ |
| 2019 8.947332 8.131136 7.968052 6.845244 4.171084 6.413981 |
| 2020 7.320804 7.501785 6.413675 3.388543 3.185676 0.950357 |
| 2021 8.206953 3.356032 2.747186 2.858031 1.018073 5.350296 |
| 2022 0.056571 3.820977 5.281980 7.199414 14.391193 7.305501 |
| 2023 6.955456 5.665194 10.228772 14.931506 17.383827 7.698504 |
|  |
| Region 24 25 |
| 2019 8.151501 2.928555 |
| 2020 8.369142 2.861033 |
| 2021 5.644463 2.299153 |
| 2022 2.549059 12.246613 |
| 2023 11.970735 14.887895 |

*Tabell 3: Procentuellt fel av LSTM-modellen för varje region och år, %*

I genomsnitt var felet per år som följer:

2019 8.85%

2020 8.19%

2021 7.55%

2022 6.23%

2023 8.45%

Det finns dock också betydande avvikelser, där det högsta felet uppgick till 19%. Detta tyder på att modellen är funktionell men kräver förbättringar.

För att förbättra modellen kan följande åtgärder övervägas:

1. Öka datasetets storlek: Genom att inkludera fler år eller ytterligare variabler, såsom socioekonomiska faktorer, kan modellen få en mer komplett bild av demografiska trender och mönster.
2. Hyperparameter-tuning: Justering av modellens hyperparametrar, som antalet noder i dolda lager och inlärningshastighet, kan leda till bättre prestanda.
3. Fler lager i nätverket: Att öka djupet på LSTM-nätverket kan hjälpa till att fånga mer komplexa mönster i datan.
4. Ensemble-metoder: Kombinera LSTM-modellen med andra modeller, som Random Forest eller XGBoost, för att öka robustheten och noggrannheten i prognoserna.
5. Regelbundenisering: Användning av metoder som Dropout eller L2-regelbundenisering kan hjälpa till att förhindra överanpassning av modellen.
6. Felsökning och analys av utvärderingsmetoder: Noggrannare utvärdering av modellens prestanda genom att analysera vilka specifika faktorer som bidrar till de stora avvikelserna kan ge insikter om hur modellen kan justeras.

Men inom ramen för detta arbete kommer vi att avstå från vidare förbättringar av modellen och de prognostiserade uppgifterna kommer att sparas i SQL-databasen i tabellen "predictions\_birth\_2024\_2029".

Denna metod säkerställer att framtida analyser kan göras baserat på nuvarande förutsägelser, även om det finns utrymme för ytterligare optimering och förbättring av prognosmodellen vid ett senare tillfälle.

För att underlätta vidare arbete kommer tabellen med de förutsagda födelsetalen, "predictions\_birth\_2024\_2029", att slås samman med de faktiska födelsetalen från SCB-statistiken, "birth\_data". Den sammanslagna tabellen kommer sedan att sparas i SQL-databasen under namnet "birth\_2007\_2029".

Denna sammanslagning gör det möjligt att ha en komplett översikt över både historiska och framtida födelsetal, vilket underlättar vidare analys och prognosarbete.

### 3.5. Beräkning av antalet elever från år 2025 till 2035

För att beräkna antalet studenter under perioden 2025–2035 använder vi tabellen med antalet födda mellan 2007 och 2029, "birth\_2007\_2029", samt de tidigare beräknade dödlighetskoefficienterna (tabell "mortality\_rate") och migrationskoefficienterna (tabell "migration\_rate"), tillsammans med *ekvation1*.

För att förenkla analysen och få fram en slutgiltig tabell i önskat format, skapar vi en tabell där analyserade år (från 2025 till 2035) anges vertikalt, medan klasserna i grundskolan och gymnasiet anges horisontellt. Vi fyller i tabellen med födelseåren för de barn som kommer att börja i respektive klass det aktuella året (tabell ”year\_klass”).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| year | F\_school | 1\_school | 2\_school | 3\_school | 4\_school | 5\_school | 6\_school | 7\_school | 8\_school | 9\_school | | 1\_gymnasium | 2\_gymnasium | 3\_\_gymnasium |
| 2025 | 2019 | 2018 | 2017 | 2016 | 2015 | 2014 | 2013 | 2012 | 2011 | | 2010 | 2009 | 2008 | 2007 |
| 2026 | 2020 | 2019 | 2018 | 2017 | 2016 | 2015 | 2014 | 2013 | 2012 | | 2011 | 2010 | 2009 | 2008 |
| 2027 | 2021 | 2020 | 2019 | 2018 | 2017 | 2016 | 2015 | 2014 | 2013 | | 2012 | 2011 | 2010 | 2009 |
| 2028 | 2022 | 2021 | 2020 | 2019 | 2018 | 2017 | 2016 | 2015 | 2014 | | 2013 | 2012 | 2011 | 2010 |
| 2029 | 2023 | 2022 | 2021 | 2020 | 2019 | 2018 | 2017 | 2016 | 2015 | | 2014 | 2013 | 2012 | 2011 |
| 2030 | 2024 | 2023 | 2022 | 2021 | 2020 | 2019 | 2018 | 2017 | 2016 | | 2015 | 2014 | 2013 | 2012 |
| 2031 | 2025 | 2024 | 2023 | 2022 | 2021 | 2020 | 2019 | 2018 | 2017 | | 2016 | 2015 | 2014 | 2013 |
| 2032 | 2026 | 2025 | 2024 | 2023 | 2022 | 2021 | 2020 | 2019 | 2018 | | 2017 | 2016 | 2015 | 2014 |
| 2033 | 2027 | 2026 | 2025 | 2024 | 2023 | 2022 | 2021 | 2020 | 2019 | | 2018 | 2017 | 2016 | 2015 |
| 2034 | 2028 | 2027 | 2026 | 2025 | 2024 | 2023 | 2022 | 2021 | 2020 | | 2019 | 2018 | 2017 | 2016 |
| 2035 | 2029 | 2028 | 2027 | 2026 | 2025 | 2024 | 2023 | 2022 | 2021 | | 2020 | 2019 | 2018 | 2017 |

*Tabell 4: Skolår, klass och födelseår för barnet.*

Vi ändrar födelseåret med antalet barn beräknat med hjälp av Python (basic användas *ekvation1*). Den resulterande tabellen (”school\_gymnasium\_results”) lagras i SQL-databas.

### 3.6. Prognostisering av kostnaden för utbildning per elev med hjälp av Linjär Regression

I detta arbete har linjär regression använts för att förutsäga framtida utbildningskostnader per elev, baserat på historiska kostnadsdata ”Kostnader för grundskola och gymnasieskola efter skolform, huvudman och kostnadsslag. År 2007 – 2022”. Genom att analysera kostnadsutvecklingen mellan 2007 och 2022 har modellen kunnat skapa prognoser för perioden 2025 till 2035. Detta har gett en enkel och effektiv metod för att uppskatta hur kostnaderna kan förändras över tid, vilket kompletterar de mer avancerade prognosmodellerna som används i studien.

En bild som visar text, linje, Graf, diagram

Automatiskt genererad beskrivning

*Figur 2. Faktiska och prognostiserade kostnader per elev i grundskolan*

En bild som visar text, diagram, linje, Graf

Automatiskt genererad beskrivning

*Figur 3. Faktiska och prognostiserade kostnader per elev i gymnasie*

Valet av linjär regression baserades på att kostnaden för utbildning per elev i grundskola och gymnasium har ökat stadigt under perioden 2007–2021. Även om det skedde en viss nedgång 2022, har vi inga data som förklarar orsaken till detta, och det finns inga tecken på att denna nedgång kommer att fortsätta. Därför antar vi att tillväxttakten kommer att förbli densamma under perioden 2025–2035. Linjär regression ger en enkel och tydlig metod för att göra en sådan prognos och förutsäga framtida kostnadsutveckling på ett tillförlitligt sätt.

De data som erhållits genom linjär regression sparas i tabellerna "grundskola\_costs\_forecast" och "gymnasieskola\_costs\_forecast" i SQL-databasen.

### 3.7. Beräkning av den totala utbildningskostnaden per region och år

Den totala kostnaden för utbildning för varje år under perioden 2025–2035, uppdelad efter varje region och år, beräknas med hjälp av Python och enligt följande ekvation:

Total\_kostnad\_grundskola\_n,m = antal\_studenter\_n,m\*kostnad\_per\_student\_n *(ekv.2)*

där n är året och m är regionen.

De erhållna data sparas i tabellerna Cost\_region\_grundskola\_forecast\_2025\_2035 och Cost\_region\_gymnasieskola\_forecast\_2025\_2035 i SQL-databasen.

# Resultat och Diskussion

### 4.1. Streamlit

Resultaten av arbetet presenteras i Streamlit genom filen "streamlit\_app.py".

En bild som visar text, skärmbild, Teckensnitt

Automatiskt genererad beskrivning

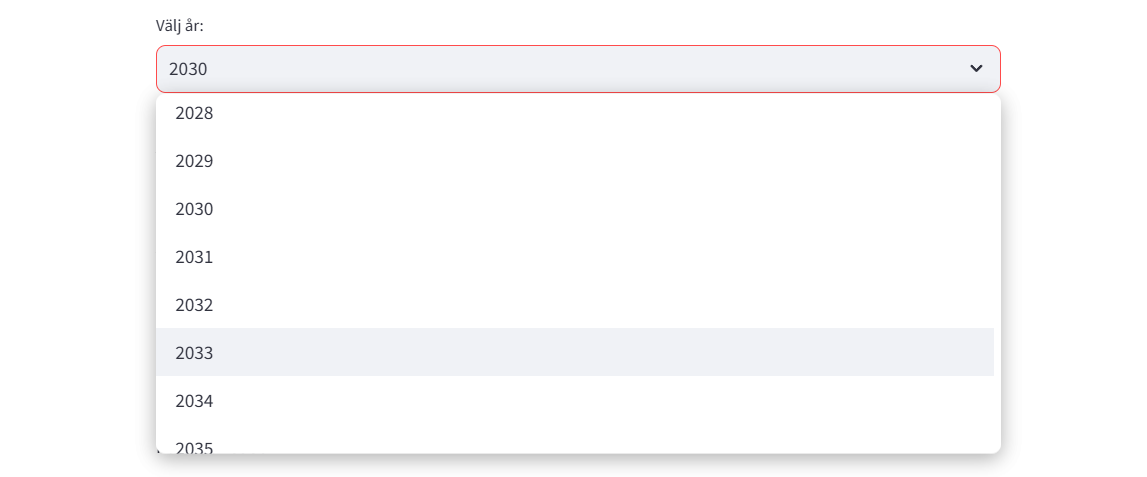


*Figur4: Utseende av Streamlit*

Från rullgardinsmeny kan användaren välja de intressanta region och år En bild som visar text, Teckensnitt, nummer, skärmbild

Automatiskt genererad beskrivning

*Figur5: Val av region*



*Figur6: Val av år*

Därefter får användaren en fullständig rapport om antalet elever, kostnader per elev och de totala utgifterna för grundskolan och för gymnasium.

En bild som visar text, skärmbild, Teckensnitt, dokument

Automatiskt genererad beskrivning

En bild som visar text, skärmbild, Teckensnitt, algebra

Automatiskt genererad beskrivning

*Figur6: Rapport i Streamlit*

### 4.2. Diskussion

Under arbetet med huvudprogrammet i filen "Projekt.ipynb" hämtas statistiska data via API, omvandlades och sparas i SQL-databas för att bli lättare att använda. Andelen dödlighet har beräknats för varje åldersgrupp (0-18 år) och andelen migration för varje åldersgrupp (0-18 år) per region. Verifieringen av dessa beräkningar visade en hög grad av tillförlitlighet, vilket gör det möjligt att använda dödlighet, andelen migration och antalet barn som var födda att beräkna antalet barn från 0 till 19 år.

Med hjälp av statistik ” Levande födda efter region och år. År 1968 - 2023” (SCB, Internet) förutspåddes LSTM-modell antalet barn som kommer att födas mellan 2024 och 2029. Modellen kunde fånga trender både per region och över tid, med en felmarginal på cirka 19 %. För praktisk användning kan dock modellen behöva finjusteras.

Med det faktiska antalet födda barn per region från 2007 till 2023 och de prognostiserade värdena från 2024 till 2029 beräknades antalet elever i varje årskurs med hjälp av dödlighets- och migrationskoefficienterna.

Kostnadsprognosen för utbildning under perioden 2025 till 2035 utfördes med hjälp av linjär regression, baserad på statistiska data från 2007 till 2022. Linjär regression är enkel att implementera, men har begränsningar när det gäller att fånga komplexa samband och förändringar över tid. För att förbättra noggrannheten kan mer avancerade modeller, som Random Forest Regressor eller Gradient Boosting, användas.

Resultaten visualiseras med hjälp av applikationen Streamlit.

# Självutvärdering

1. Utmaningar du haft under arbetet samt hur du hanterat dem.

Det var svårt att välja rätt modell och sedan justera parametrarna för att minska felet. Jag har provat många olika modeller

1. Vilket betyg du anser att du skall ha och varför.

G eller VG

# Källförteckning:

1. Nelson D., ” Vad är RNN och LSTM i Deep Learning?”, <https://www.unite.ai/sv/vad-%C3%A4r-rnns-och-lstms-i-djupinl%C3%A4rning/> (18.10.2024)
2. Prgomet A., GitHub, Recurrent Neural Networks (RNN), <https://github.com/AntonioPrgomet/djupinlaerning/blob/main/05_rnn.pptx>
3. Géron, A, Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow, 2019, s.516
4. IBM ” What is linear regression?”, <https://www.ibm.com/topics/linear-regression#:~:text=register%20for%20TechXchange-,What%20is%20linear%20regression%3F,is%20called%20the%20independent%20variable> (18.10.2024)
5. GeeksforGeeks , Linjär regression i maskininlärning, <https://www.geeksforgeeks.org/ml-linear-regression/>, (21.10.2024)

6. SCB, Levande födda efter region och år. År 1968 - 2023 , <https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101H/FoddaK/>

7. SCB, Folkmängd 0-19 år efter region, ålder och år, https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START\_\_BE\_\_BE0101\_\_BE0101A/BefolkningNy/

8. SCB, Döda efter region, ålder och år , <https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101I/DodaFodelsearK/>

9. SCB, Flyttningsöverskott efter region, ålder och år, <https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101J/Flyttningar97/>

10. SCB, Kostnader för grundskola och gymnasieskola efter skolform, huvudman och kostnadsslag. År 2007 - 2022, https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START\_\_UF\_\_UF0514/UtbKostGrundGym/