Runskrift Projekt



Jesper Kokkonen

EC Utbildning

2024–10

# Abstract

This project applies machine learning to analyze and classify runestones. The data is then integrated into Power BI to create interactive visualizations, including a map displaying the geographical distribution of runestones. By leveraging these technologies, we gain valuable insights into the history and patterns of runestone inscriptions

**Skapas automatiskt i Word genom att gå till Referenser > Innehållsförteckning.**

Innehållsförteckning

[Abstract 2](#_Toc181362781)

[1 Inledning 1](#_Toc181362782)

[1.1 GitHub 1](#_Toc181362783)

[2 Teori 3](#_Toc181362784)

[2.1 GitHub 3](#_Toc181362785)

[2.2 Klassificeringsproblem 3](#_Toc181362786)

[2.2.1 Logistisk Regression 3](#_Toc181362787)

[2.2.2 Naive Bayes 3](#_Toc181362788)

[2.2.3 Support Vector Machine (SVM) 3](#_Toc181362789)

[2.2.4 K-Nearest Neighbors (KNN) 4](#_Toc181362790)

[2.2.5 Decision Tree 4](#_Toc181362791)

[2.2.6 Random Forest Classifier (RFC) 4](#_Toc181362792)

[2.2.7 Word2Vec 4](#_Toc181362793)

[2.3 Modellutvärdering 4](#_Toc181362794)

[2.3.1 Confusion Matrix 4](#_Toc181362795)

[2.3.2 Mätvärden 5](#_Toc181362796)

[3 Metod 6](#_Toc181362797)

[3.1 Machine Learning 1 6](#_Toc181362798)

[3.2 Machine Learning 2 7](#_Toc181362799)

[3.3 Konvertering av SWEREF99 och WGS84 7](#_Toc181362800)

[3.4 Agil arbetsmetodik 8](#_Toc181362801)

[4 Resultat och Diskussion 9](#_Toc181362802)

[4.1 Machine Learning 1 9](#_Toc181362803)

[4.2 Machine Learning 2 11](#_Toc181362804)

[4.3 Power BI 12](#_Toc181362805)

[4.3.1 Analys av materialdifferentiering i runstenar över tidsperioder och regioner 14](#_Toc181362807)

[5 Slutsatser 15](#_Toc181362808)

[6 Självutvärdering 16](#_Toc181362809)

[Källförteckning 17](#_Toc181362810)

# Inledning

Projektet syftar till att utveckla och implementera två maskininlärningsmodeller för att klassificera och analysera runstenar. Modellerna kommer att tränas på två olika databaser av runinskriptioner. Resultaten kommer att integreras i en Power BI-dashboard, inklusive en geospatial analys för att visualisera den geografiska fördelningen av runstenar. Genom att kombinera maskininlärning och datavisualisering möjliggör detta projekt en djupgående analys av runstenarnas lingvistiska och kulturella egenskaper.

1) Finns det betydande skillnader mellan runinskrifterna för att möjliggöra klassificering i olika tidsperioder?

2) Är det möjligt att fullfölja dateringen av runstenarna genom en annan strategi än den visuella observationen av individuella stildrag eller tillsammans med den?

3) Går det att kontrollera dateringen genom kartstudie av olika sekvenser av ristningar och deras geografiska spridning?

## GitHub

I vår GitHub-repo hittar du en omfattande README-fil som fungerar som din guide till projektet. Vi börjar med att ge dig en tydlig överblick över vad projektet syftar till.

Steg för steg leder vi dig genom projektets arbetsflöde, så att du enkelt kan förstå hur alla delar hänger ihop. Vi har också tagit fram en detaljerad installationsguide som tydliggör steg för steg installationsprocessen. Här hittar du all nödvändig information om vilka programvaror och bibliotek du behöver

Med README-filen som din guide kommer du snabbt att kunna komma i gång med att använda och utforska projektet på egen hand. Några av fördelarna med GitHub är att det blir lättare att samarbeta med flera olika utvecklare som arbetar på samma projekt. GitHub gör det enkelt att se vem som har gjort vilka ändringar och att slå samman olika versioner av kod.

A screenshot of a project

Description automatically generated

Kontrollen av vårt projekt på GitHub var enkelt, tack vare GitHub spårning av kodändringar över tid. Det möjlig gjorde för var och en att kunna se vid varje commit vad som laddades upp på vår gemensamma GitHub vilket gjorde det enkelt att förstå och återse utvecklingen av projektet.

Om ett problem uppstod eller någon oavsiktligt ändrade vår GitHub så kunde vi snabbt att återgå till en tidigare version av projektet.

GitHub har en Project board feature där vi skapade en Kanban för att kunna spåra framsteg på ett snabbt sätt. Där organiserade vi olika uppgifter och funktioner i tre olika stadier som ”In Progress”,

”Ready” och ”Done”, så att gruppen kunde visualisera projektets övergripande status.

A screenshot of a computer

Description automatically generated

# Teori

## GitHub

GitHub låter dig skapa, lagra, ändra, slå samman och samarbeta på filer eller kod. Alla medlemmar i ett team kan komma åt GitHub-förvaret (tänk på detta som en mapp för filer) och se den senaste versionen i realtid. Sedan kan de göra ändringar som de andra samarbetspartnerna också ser. GitHub låter också användare göra förfrågningar till varandra och internt diskutera iterationerna längs vägen. GitHub har även kallats för en "a social coding platform" eftersom den bjuder in människor att koordinera, dela och samarbeta med kod över distribuerade och asynkrona miljöer

## Klassificeringsproblem

Klassificeringsproblem går ut på att placera data i en av flera fördefinierade kategorier. Modeller som används går ut på att generalisera och korrekt klassificera ny, osedd data baserat på relationer och mönster som identifierats i träningsdatan under träningsfasen. Det finns binära klassificerare som skiljer mellan två klasser, samt multinomialklassificerare som skilja mellan fler än två klasser. (Géron, 2019, s. 100).

### Logistisk Regression

Logistisk regression är en metod särskilt lämplig för att hantera binära klassificeringsproblem där målet är att förutsäga sannolikheten för att en observation tillhör en viss kategori (oftast betecknad som 0 eller 1). Kan tillämpas inom t.ex. forskning om kundbeteenden för att bedöma sannolikheten för att en kund svarar positivt eller negativt på en kampanj. (James, m.fl., 2021, s. 133).

### Naive Bayes

Naive Bayes utgår ifrån att variablerna är oberoende av varandra och att alla variablerna bidrar lika mycket till resultatet. Detta antagande stämmer inte för de flesta verkliga situationer, men det förenklar de numeriska beräkningarna och gör problemet lättare hanteringsbart. (Shukla, 2024).

### Support Vector Machine (SVM)

SVM är lämpad för att hantera klassificeringsproblem, då dess mål är att separera olika klasser så precist som möjligt, genom att hitta en linje eller hyperplan som har den minsta möjliga avvikelsen från datapunkterna. Med SVM har man viss flexibilitet att definiera hur mycket fel som är acceptabelt i modellen, genom att justera parametern C. (Bajallan, 2022, s. 10).

### K-Nearest Neighbors (KNN)

Med KNN lagrar man alla tillgängliga värden och klassificerar varje ny data baserat på dess likhet med datapunkten som ligger närmast. Tanken bakom modellen är att punkter som ligger nära varandra brukar tillhöra samma kategori. (Bajallan, 2022, s. 16).

### Decision Tree

Beslutsträd är en övervakad algoritm som bygger strukturer som liknar träd, med s.k. beslutsnoder och lövnoder, där varje nod förgrenas till ett antal nya noder beroende på antalet möjliga utfall. De noder som inte själv förgrenas kallas för ”löv”. Beslutsträd är en serie beslut som tas utifrån given data i en bestämd följdordning för att nå ett specifikt resultat. Medan algoritmen är ganska lätt att använda, är ett enda träd oftast inte tillräckligt för att ge effektiva resultat. (Sharma, 2024).

### Random Forest Classifier (RFC)

RFC kombinerar förutsägelserna från flera individuella beslutsträd för att generera det slutliga resultatet. Detta är en slumpartad teknik som kombinerar resultatet av flera individuella modeller och fattar det slutliga beslutet baserat på majoriteten av utfallen. RFC kan på så vis förbättra modellens övergripande prestanda och robusthet jämfört med enskilda beslutsträd. (Sharma, 2024).

### Word2Vec

Word2Vec is a neural network-based model that groups words with similar contexts by analyzing surrounding words in sentences or documents. It creates word embeddings, where similar words are represented by vectors positioned close together in a high-dimensional space. Each word is mapped to a unique vector based on its semantic meaning and relation to other words. Word2Vec is foundational in natural language processing (NLP) and powers applications like machine translation, speech recognition, and AI chatbots.

## Modellutvärdering

### Confusion Matrix

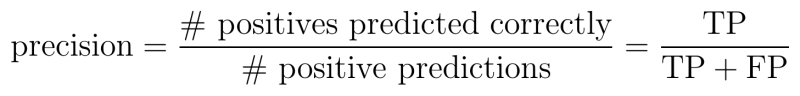
Förvirringsmatrisen (confusion matrix) är en tabell som används för att utvärdera prestandan för en klassificeringsmodell, genom att tillhandahålla en sammanfattning av de förutsägelser som modellen gör. Förvirringsmatrisen är en tabell som visar fyra värden:

• True positive (TP): Datapunkter som vi förutspådde positiva som faktiskt är positiva.

• False positive (FP): Datapunkter som vi förutspådde positiva som faktiskt är negativa.

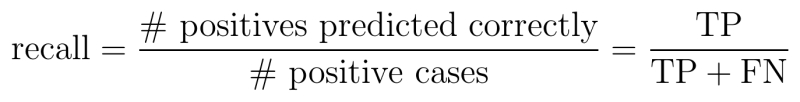
• False negative (FN): Datapunkter som vi förutspådde negativa som faktiskt är positiva.

• True negative (TN): Datapunkter som vi förutspådde negativa som faktiskt är negativa.  
  
Den första och sista är de datapunkter som förutspåddes korrekt och den andra och tredje är de datapunkter som förutspåddes felaktigt. Vi kan använda de fyra värdena i förvirringsmatrisen (TP, TN, FP, FN) för att beräkna olika mätvärden som noggrannhet, precision, träffsäkerhet och F-mått. (Narkhede, 2018)



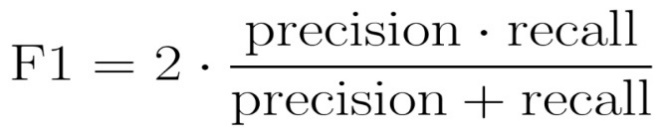
Figur 1. Precision = Av alla datapunkter som vi förutspått som positiva, så många är faktiskt positiva.

Träffsäkerhet (Recall)är andelen positiva fall som modellen förutsäger korrekt. (Narkhede, 2018)



Figur 2. Träffsäkerhet = Av alla datapunkter med positivt värde, så här många förutspåddes som positiva (alltså korrekt).

F1-mått (F1 score) är ett genomsnittsvärde av precision och träffsäkerhet, en kompromiss när vi inte kan välja mellan två modeller då den ena har högre precision och den andra högre träffsäkerhet. (Géron, 2019, s. 92).



Figur 4. Den matematiska definitionen av F1-måttet där man multiplicerar produkten av precisionen och träffsäkerheten

med 2, och delar resultatet med summan av precisionen och träffsäkerheten.

### Mätvärden

Noggrannhet (Accuracy) avser procentandelen förutsägelser som är korrekta. Om vi t.ex. har 100 datapunkter och förutsäger 70 av dem korrekt är noggrannheten 70/100 = 0.7 = 70 %. Noggrannhet är ett bra mått om våra klasser är jämnt fördelade, men kan vara missvisande vid obalans mellan klasserna. (Géron, 2019, s. 90). Därför kan det med fördel kompletteras med precision, recall och F score för en mer detaljerad bild.

Precision är procentandelen av modellens positiva förutsägelser som är korrekta. (Narkhede, 2018)

# Metod

## Machine Learning 1

Syftet har varit att förutsäga saknade värden i kolumnen "Period/Datering" baserat på värden i de relaterade kolumnerna "Kommun", "Koordinater", "Materialtyp" och "Föremål". Uppgiften har hanterats som ett klassificeringsproblem, då vi vill förutsäga ett kategoriskt utfall (värdet i "Period/Datering") utifrån andra variabler. Därför har Logistic Regression eller Support Vector Classifier (SVC) valts som lämpliga modeller.

Första steget innebar att förbereda och formatera data, så att variablerna hade lämpligt format som text, nummer eller tid. Vi använde två metoder för datarensning. Den första kategoriserade årtal och kortare perioder till större tidskategorier (se bilaga - - - - -), vilket gav totalt 20 kategorier (15 kända och 5 okända) där varje kategori representerade en tidsperiod.

Den andra metoden använde Regex (Regular Expression), ett verktyg för att identifiera och extrahera textmönster, för att hitta årtalsintervall (t.ex. 1100-1150 eller 100-200). Funktionen clean\_date identifierade första förekomsten av mönstret \d{3,4}-\d{3,4} i texten och ignorerade andra tecken som kommentarer och bokstäver. Resultaten sparades sedan i en ny kolumn.

Det visade sig att regex-mönstret inte matchade all data. Mönstret \d{3,4}-\d{3,4} söker specifikt efter intervall som "1100-1200", men databasen innehåller även andra format, såsom "900-t" eller "efter 1050", som inte matchar detta. För att göra funktionen mer flexibel uppdaterades regex för att kunna hantera fler numeriska format, inklusive:

* Enstaka årtal (t.ex. 900)
* Årsintervall (t.ex. 1100-1150)
* År med tillägg av bokstäver eller symboler (t.ex. 900-t eller 1100-t)

Funktionen uppdaterades med ett mer omfattande regex

Det uppdaterade regex-mönstret \d{3,4}(-\d{3,4})? matchade både enskilda årtal (t.ex. 1100) och intervall (t.ex. 1100–1150), vilket gjorde det möjligt att samla in korrekt årsdata och ignorera bokstäver och symboler. Med de tvättade data kunde vi nu förutsäga perioder för instanser utan värden, baserat på data från instanser där perioder är angivna. Eftersom vi saknar exakta datering för vissa runstenar, jämförde vi prediktionerna från två olika metoder för datarensning och kontrollerade mot ristarnas aktiva perioder i kolumnen "Ristare".

För att välja den bästa modellen analyserade vi noggrannhet, precision, träffsäkerhet och F1-poäng från klassificeringsrapporter för Logistic Regression, SVC och RandomForestClassifier, där data delades med train\_test\_split. Under testningen uppstod dock varningen "UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples", vilket är vanligt när vissa klasser är underrepresenterade i tränings- eller testuppsättningen, och modellen kan då inte klassificera dem korrekt.

Varningen hanterades genom att kontrollera klassfördelningen och justera zero\_division-parametern. Det visade sig att kolumnen ”Föremål” var mycket obalanserad, där vissa klasser hade många instanser (t.ex. klass 785 med 172 prover) och andra hade få (t.ex. klasser med endast 1 prov). Denna obalans påverkade modellernas prestanda för underrepresenterade klasser. Samma kontroll utfördes för övriga variabler.

Vid fortsatt testning användes modeller med klassviktning (t.ex. RandomForestClassifier(class\_weight='balanced', random\_state=42)) för att automatiskt justera vikter efter klassfördelningen. Dock kvarstod obalansen i klasserna och påverkade prediktionerna. Vi observerade detta när vi jämförde predikterade perioder med ristarnas verksamma perioder, vilket redovisas i nästa avsnitt.

## Machine Learning 2

Det första steget i projektet var att utveckla ett skript för att automatisera insamlingen av data om runinskrifter från Wikipedias API. Målet var att hämta tillgänglig information från Wikipedias sidor under kategorin "Upplands runinskrifter". Skriptet navigera genom kategorin och samlar in detaljerad information om varje runstens placering, historiska betydelse, tolkningar och andra beskrivande data som fanns tillgängliga.

Skriptet började med att anropa Wikipedias API för att få en lista över alla sidor som ingår i kategorin "Upplands runinskrifter". För varje sida i listan hämtade skriptet sedan innehållet genom att använda API-förfrågningar, där den utvalda informationen extraherades och bearbetades. Detta inkluderade att strukturera och rensa upp data för att säkerställa att den var konsekvent och klar för vidare analys. När alla nödvändiga data hade samlats in och bearbetats lagrades den i en CSV-fil som döptes till *Uppland.csv*. CSV-formatet valdes för sin enkelhet och kompatibilitet med många analysverktyg, vilket gör filen enkel att importera och analysera.

Nästa steg är att använda Word2Vec för att skapa semantiska ordinbäddningar av runtexten. Denna metod är särskilt användbar för att undersöka språkliga mönster och samband i inskriptionerna, som kan avslöja information om inskriptionsstilar, perioder eller regionala drag.

## Konvertering av SWEREF99 och WGS84

Scandinavian Runic-text Database använder det svenska koordinatsystemet SWEREF99, vilket begränsar användningen till Sverige. Vår databas lagrar bara den nordliga (N) koordinaten, medan den östliga (E) koordinaten saknas. PowerBI använder dock WGS84-koordinatsystemet. För att visualisera data i PowerBI måste vi konvertera SWEREF99-koordinater till WGS84. Flera bibliotek kan användas för denna konvertering:

.Net - [MightyLittleGeodesy](https://github.com/bjornsallarp/MightyLittleGeodesy)

Java - [CoordinateTransformationLibrary](https://github.com/goober/coordinate-transformation-library)

PHP - [CoordinateTransformationLibrary](https://github.com/david-xelera/CoordinateTransformationLibrary)

JavaScript - [latlong.mellifica.se](http://latlong.mellifica.se/)

Python - pyproj, Transformer

Vi testade Transformer-klassen i pyproj-biblioteket för att konvertera mellan koordinatsystem. Vi har framgångsrikt omvandlat den nordliga (N)-koordinaten till longitud. Men på grund av saknade Easting (E) koordinater måste vi hitta alternativa metoder för att komplettera data.

params = {

    'action': 'query',

    'prop': 'coordinates',

    'titles': page\_title,

    'format': 'json'

}

Vi utforskade också MediaWiki API för att hämta latitud- och longituddata för specifika platser. Även om detta tillvägagångssätt fungerar, kräver det att var och en av de 11672 posterna behandlas individuellt, vilket kan vara beräkningsmässigt dyrt för stora datamängder.

Vi försökte kartlägga latitud till kolumnen "Koordinater" baserat på kolumnen "Kommun". Denna metod, samtidigt som den tar itu med problemet med saknade data, saknar precision. Den kan lokalisera gravstenen i rätt kommun men kan inte fastställa dess exakta plats, vilket begränsar noggrannheten av visualiseringen i PowerBI.

## Agil arbetsmetodik

När det kommer till Agila arbetsmetodik så tycker jag att vi har följt mycket av det som står i agile manifesto. Vi utvecklade två någorlunda fungerande machine learning model väldigt tidigt som vi sedan byggde på och förbättrade med tiden. I manifestet så nämner dem att man ska mötas ansikte mot ansikte det var inte vi kunde göra men vi hade dagliga möten på teams där vi kunde prata om hur det går och om någon har stött på problem.

# Resultat och Diskussion

## Machine Learning 1

Detta är de nya värdena, avseende Period/Datering efter att datan tvättats med Regex:

**A white screen with black text

Description automatically generated**

Samtliga värden (text, bokstäver, symboler, mm) förutom de numeriska har tagits bort, vilket gör att värden som avser samma tidsperiod, men skrivits på olika sätt, blir identifierade som likadana. Resultatet blev att de unika värdena minskat från 587 till 241.

Nedan ser vi de unika värdena i kolumnen "Period/Datering" efter att datan indelats i kategorier:

numeric\_values = [

'folkvandringstid 160–375',

'högmedeltid 1150–1300',

'mellanvendeltid 600–700',

'modern tid 1500–1700',

'modern tid 1700–1800',

'modern tid 1800–1900',

'sen vikingatid 1050-1200',

'senmedeltid 1300–1500',

'sent folkvandringstid 375–500',

'sent vendeltid 700–800',

'tidig medeltid 1000–1150',

'tidig vendeltid 500–600',

'tidig vikingatiden 700-800',

'vikingatiden 800-1050',

'vikingatiden 800–1050'

]

non\_numeric\_values = [

'allmän/osäker sentida',

'oklar/allmän vendeltid',

'okänd',

'osäker/allmän medeltida',

'unknown'

]

Så här såg värdena ut efter att prediktionerna gjorts (exemplet visar data som tvättats med första metoden, dvs. blivit indelad i kategorier):

Period/Datering Predicted\_Period/Datering

0 allmän/osäker sentida vikingatiden 800–1050

1 allmän/osäker sentida senmedeltid 1300–1500

2 allmän/osäker sentida vikingatiden 800-1050

3 allmän/osäker sentida vikingatiden 800-1050

4 allmän/osäker sentida vikingatiden 800-1050

... ... ... ... ...

11666 vikingatiden 800–1050 NaN

11667 vikingatiden 800–1050 NaN

11668 vikingatiden 800–1050 NaN

11669 vikingatiden 800–1050 NaN

11670 vikingatiden 800–1050 NaN

Rader i tabellen som saknar ett numeriskt värde för kolumnen ”Period/Datering” har ersatts med predikterade värden. För raderna där det finns ett numeriskt värde har givetvis inga prediktioner gjorts, utan dessa rader har utgjort träningsdatan.

Så här ser resultatet ut med modellen som visats sig bäst (RFC):

Totala matchningar: 6435

Korrekt matchningar (True Positives): 5745

Felaktiga matchningar (False Positives): 690

Felaktiga icke-matchningar (False Negatives): 122

Accuracy: 89.28%

Precision: 89.28%

Recall: 97.92%

F1 Score: 93.40%

Resultaten visar totalt 6435 identifierade matchningar mellan värdena i de två kolumnerna. Av dessa matchningar är 5745 korrekta (True Positives), vilket innebär att värdena i de två kolumnerna stämmer överens med förväntade perioder enligt indelningen. Felaktiga matchningar, där data felaktigt bedömts som överensstämmande men inte matchar korrekt, uppgår till 690 (False Positives). Antalet missade korrekta matchningar (False Negatives) är 122, vilket betyder att dessa matchningar inte identifierades.

Noggrannheten (Accuracy) ligger på 89,28%, vilket innebär att nästan 90 procent av alla matchningsförsök har varit korrekta, vilket tyder på hög precision och få felaktiga matchningar. Träffsäkerheten (Recall) ligger på nästan 98%, vilket visar att metoden effektivt identifierar verkliga matchningar även om vissa felaktigt bedöms som överensstämmande. F1-värdet, som kombinerar precision och recall, är 93,4%, vilket bekräftar metodens övergripande tillförlitlighet att korrekt identifiera matchningar och minimera fel.

Vid en jämförelse mellan våra predikterade perioder och ristarnas verksamma perioder presterar modellen mycket bra på majoritetsklasser (ristare med många instanser) men har svårigheter med minoritetsklasser (ristare som förekommer i få, ofta bara enstaka instanser).

## Machine Learning 2

För analys av personer som skapat runskrifter under sitt liv har vi skapat en KNN-regressionsmodell för att göra förutsägelser på specifika runinskriptioner. Den börjar med att ladda modellen och datauppsättningen, rensar data, tillämpar ett filter för specifika inskrivare och gör förutsägelser baserat på dessa filtrerade poster.

För att möjliggöra en fördjupad analys av individuella runinskrivare har vi implementerat en

K-Nearest Neighbors (KNN)-regressionsmodell. Modellen tränas på vår CSV-fil bestående av runinskriptioner, där varje datapunkt representerar en unik inskription och dess associerade metadata (t.ex. datering, geografisk placering, stilistiska drag).

För att visualisera fördelningen av runinskriptioner har vi en karta. Varje runsten representeras av en markör och ett årtal. Genom att filtrera efter specifika inskrivare eller tidsperioder kan användaren visualisera geografiska kluster för runmästare.

A blue and brown background with lines and numbers

Description automatically generated

Kartan visualiserar en runmästares färd under dess arbete, illustrerat genom en animerad linje som binder samman de olika platser där denne har skapat runinskrifter. Denna visuella representation ger oss en inblick i runmästarens rörelsemönster.

A table with numbers and numbers

Description automatically generated

Bilden visar runinskriptioner, där varje inskription är markerad med sin geografiska position (latitud och longitud), årtalet för inskriptionen och dess unika signum

## Power BI

|  |  |
| --- | --- |
| 1730129111(1) | 1730129159(1) |

Figur 1: (a) Trend Analysis of Föremål Over Time and Location. (b) Trend Analysis of Stilgruppering and Signum: Temporal, Geographical, and Spatial Distribution

Figur 1(a) visar att gravstensanvändningen ökade markant under vikinga- och medeltiden, med 62 unika Föremål under vikingatiden och 78 under medeltiden, jämfört med endast cirka 10 under tidigare perioder. Tidiga Föremål var enklare och mer jämnt fördelade, främst i Svealand, men senare koncentrerades gravstenar till Uppland, vilket tydde på bildandet av större centraliserade samhällen. Dessutom var runstenar fler än gravstenar, med sten som det dominerande materialet över trä eller metall. Som Alix Thoeming (University of Sydney) noterar representerar ökningen av runstenar ett kulturellt uttryck för tro och identitet som förminskats av kristnandet [1].

* A: Står ofta för "Äldre järnåldern" (Tidig Järnålder).
* U: Representerar den typiska "Vikingatiden".
* V: Används vanligtvis för "Vendeltid" (Vendelperioden, som föregår vikingatiden i Skandinavien).
* M: Indikerar vanligtvis "Medeltiden".
* Sentia: representerar perioden mellan 1500 och 1900.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| A stone with a drawing on it  Description automatically generated  Pr 1 | | A stone with a cross carved into it  Description automatically generated  Pr 2 | | A stone with a carved figure  Description automatically generated with medium confidence  Pr 3 | A stone with celtic designs  Description automatically generated  Pr 4 | |
| A stone with celtic symbols  Description automatically generated  Pr 5 | A stone with red writing on it  Description automatically generated  RAK | | A stone with a snake carved into it  Description automatically generated  Fp | | | A stone with a celtic knot  Description automatically generated  KB |

# *Figur 2: Runestone styles from wikipedia [2]*

Pr1–Pr6 hänvisar till faser av furstliga eller proto-runiska stilar (t.ex. Ljung- eller Gräslundstilar), som visar konstnärliga och typologiska förändringar i runstensdesign. Denna serie avslutas med djurhuvuden i profil. Fp, från c. 1010–1050, har runband som slutar med djurhuvuden sett uppifrån. KB betecknar stilen Kongliga Biblioteket, kopplad till samlingar i Kungliga biblioteket. RAK, den äldsta stilen (980–1015 e.Kr.), saknar drakhuvuden, med runband som slutar i raka linjer. "sid" syftar på sten sidor med inskriptioner eller mönster. Se figur 2 för illustrationer [2].

Figur 1(b) analyserar klassificeringen av runstensdesigner efter tid och plats, och avslöjar att RAK-runstenar är de vanligaste, med 558 exempel som visar deras utbredda användning under en längre period. Pr4 följer med 427 runstenar, vilket indikerar bred representation över platser, huvudsakligen från ca. 1060–1100. KB är den mest sällsynta stilen, med endast 48 runstenar, uteslutande från vikingatiden och till stor del finns i Uppsala, Nyköping och Eskilstuna, vilket tyder på ett unikt regionalt inflytande.

Pr4-runstenar, främst vikingatiden, är mer spridda än KB och förekommer i områden som Uppsala, Gotland och Stockholm, vilket indikerar en bredare kulturell räckvidd. Däremot sträcker sig RAK från Vendel till medeltiden, med platser som Sigtuna, Lund och Norrköping. Denna breda fördelning antyder RAK:s anpassningsförmåga eller bestående symboliska värde över regioner och epoker. Sammantaget belyser dessa mönster hur runstensstilar speglar både tidsmässiga och regionala förändringar i skandinaviska konstnärliga och kulturella uttryck.

|  |  |
| --- | --- |
| A screenshot of a data presentation  Description automatically generated | A screenshot of a graph  Description automatically generated |
| (a) | (b) |

*Figur 3: (a) Tombstone Material Statistics: Trends by Time Period, Geographic Location, and Intended Purpose. (b) Ristare and korsform Statistics: Trends by Time Period, Geographic Location, and Intended Purpose.*

### Analys av materialdifferentiering i runstenar över tidsperioder och regioner

Mångfalden av material i runstenar ger insikt i de sociokulturella och miljömässiga influenserna som formar runinskrifter. Figur 3(a) visar sten som det primära materialet under vikinga-, vendel- och medeltiden, vilket återspeglar dess hållbarhet och rituella betydelse. Stenrunstenar finns främst i södra Sverige, vilket tyder på ett delat medium för kulturella uttryck över regioner.

Under medeltiden blev trä populärt, med 2 748 inspelade exempel, troligen på grund av ökad tillgänglighet och kulturella förändringar, särskilt i Norges Bergen och Trondheim, där trärunstenar är koncentrerade. Metall, även om det är sällsynt, förekommer främst på Gotland, ett rikt handelsnav, vilket tyder på högstatusanvändning. Ben-, horn- och gipsrunstenar är sällsynta och ofta kopplade till specifika regioner som Sigtuna, Lund och Oslo, förknippade med politisk eller religiös betydelse.

Sammantaget visar materialval och geografisk spridning en blandning av kulturella, resursbaserade och sociala faktorer. Sten och metall tyder på kontinuitet i kulturell vördnad, medan trä och ben indikerar anpassning till lokala resurser. Materialvalet i sig kan förmedla kulturell betydelse, där sten antyder beständighet i Sverige, och trä eller metall betecknar distinkta sociala betydelser baserat på plats och status.

Figur 3(b) avslöjar en rik mångfald av runstensristare och stilar, med över 7 000 hantverkare som bidrar till denna kulturella tradition och visar en betydande kommunal och ceremoniell roll för snidning. Bara under vikingatiden var 3 612 ristare aktiva, vilket underströk vikten av runstenar för att dokumentera människor, klaner och händelser från den eran.

Runstenar i RAK-stil, tillverkade av 558 snidare, är koncentrerade till Mjölby, Bornholm, Norrköping och Linköping, vilket speglar traditionella eller tidigare designpreferenser. Fp-stil, skapad av 216 ristare, förekommer främst i Enköping, Strängnäs och Nyköping, medan runstenar i Pr4-stil, tillverkade av 427 ristare, främst finns i Uppsala, Enköping och Sigtuna. Pr4:s mer genomarbetade design, kopplad till medeltiden, kan spegla avancerat hantverk och högre kulturell eller religiös betydelse, särskilt i Uppsala, ett framstående religiöst centrum.

Sammanfattningsvis illustrerar ristarnas stilar och distribution en levande, regionalt influerad tradition, där varje stil återspeglar unika lokala värderingar, kulturella prioriteringar och sociala influenser.

# Slutsatser

Machine Learning 1: Analysen demonstrerar att den föreslagna metoden uppvisar hög träffsäkerhet och precision vid identifiering av matchningar, vilket bekräftas av det höga F1-värdet. Detta indikerar en balanserad förmåga att både detektera relevanta matchningar och undvika false positiv. Genom att optimera periodmappningen och implementera ytterligare, mer stringenta matchningskriterier, kan vi potentiellt reducera antalet false positiv och false negativ, särskilt för minoritetsklasser, och därmed ytterligare höja metodens tillförlitlighet.

Machine Learning 2: Genom att använda modellen och kartvisualisering kan vi se olika runmästare som arbetar i olika geografiska områden. Vilket kan indikera om en runmästare var verksam i en viss region eller om deras arbete spreds över ett större geografiskt område. Med hjälp av modellen kan vi identifiera om olika trender uppstår. Exempel på en trend som kan vissas är stilistiska förändringar på runor som är kopplade till en specifik tidsperiod, detta kan antyda på kulturella skillnader och influenser under olika tidsperioder.

Interaktiv Karta: Utvecklingen av runstensändamål, material och snidningsstilar speglar hur svenska och norska samhällen balanserade tradition och religiöst uttryck med anpassningsförmåga. Materialval i linje med ceremoniella roller och symbolisk betydelse, medan variationen i snideristilar över hela Sverige illustrerar förändrade kulturella prioriteringar och influenser över tid. Sammantaget understryker dessa element runstenar som djupa kulturella artefakter, som förkroppsligar regionala identiteter, sociala värden och ett bestående arv.

# Självutvärdering

1. Utmaningar du haft under arbetet samt hur du hanterat dem.   
   Jag tyckte inte att vi hade några större utmaningar under vårt arbete. Eftersom vi möttes dagligen så kunde vi lösa utmaningar om dem hade uppstått.
2. Vilket betyg du anser att du skall ha och varför.   
   G
3. Något du vill lyfta fram till Antonio?

Nej

# Källförteckning

[1] What is GitHub and Why Should You Use It?

<https://www.coursera.org/articles/what-is-git>

[2] Bajallan, R. (2022). *A comparative evaluation of machine learning models for engagement classification during presentations.* Stockholm: KTH.

<https://www.diva-portal.org/smash/get/diva2:1732944/FULLTEXT01.pdf>

[3] Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow; concepts, tools, and techniques to build intelligent systems.* Sebastopol: O'Reilly.

[4] James, G., Witten, D., Hastie, T., Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R.* Switzerland: Springer. <https://link.springer.com/book/10.1007/978-1-4614-7138-7>

[5] Narkhede, S. (2018) I: *Towards Data Science:* ***Understanding Confusion Matrix.*** [*https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62*](https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62)

[6] Sharma, A. (2024). I: Analytics Vidhya: *Random Forest vs Decision Tree - Which is Right for You?* <https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/#:~:text=A%20decision%20tree%20is%20more,%2C%20healthcare%2C%20and%20deep%20learning>.

[7] Shukla, P. (2024). I: Analytics Vidhya: *Naive Bayes Algorithms: A Complete Guide for Beginners.* <https://www.analyticsvidhya.com/blog/2023/01/naive-bayes-algorithms-a-complete-guide-for-beginners/>

[9] What is Word2Vec and How Does It Work?

<https://swimm.io/learn/large-language-models/what-is-word2vec-and-how-does-it-work>

[10] Alix Thoeming, “HERE SHALL THESE STONES STAND, REDDENED WITH RUNES

EXPLORING INTERCONNECTIVITY AND SIMILARITY IN THE RUNE STONES OF 10TH-12TH CENTURY SWEDEN", University of Sydney (2013).

[11] <https://en.wikipedia.org/wiki/Runestone_styles>