

INLEDNING

Bakgrundsbeskrivning	<p>Vi har nu satt in foten i AI-världen, det är dags att öva på att skapa några grundläggande modeller. Följ stegen vi har gjort under lektionstid och det borde inte vara särskilt problematiskt.</p> <p>Telco AB är ett fiktivt företag som har en utmaning med kundbortfall. Man vill identifiera de kunder som riskerar att falla bort. Dessa kunder kan genom riktade erbjudanden lockas att stanna kvar.</p> <p>Själva datan är baserad på en publik Kaggle-tävling, men något modifierad. Den beskrivs här: https://www.kaggle.com/datasets/blastchar/telco-customer-churn</p>
Varför ska ni utföra detta arbete?	<p>Syftet är att visa hur man kan använda AI för att lösa ett verklighetstroget problem, genom att undersöka vilken modell som är bäst lämpad för denna uppgift. Ni skall få öva på både regression och klassificering.</p>
Vad ska ni leverera?	<p>Uppgiften lämnas in genom en eller flera Jupyter notebook (.ipynb) filer, där stegen beskrivs och där all kod, resultat, text och eventuella grafer finns samlade. Du skall använda dig av Markup Language för att beskriva delarna i rapporten. Varje del ska vara tydligt markerad med en titel, med betoning på presentation.</p>

ER PROJEKTUPPGIFT

Vad ska ni göra?	<p>Ni skall köra Machine Learning på två stycken dataset, den ena med regression, den andra med klassificering.</p> <p>För regression skall ni hitta ett dataset på internet där ni tar bort en kolumn som innehåller ett numeriskt värde och sedan använder regression för att räkna ut det. Använd flera Machine Learning modeller, minst 2, och jämför vilken som presterade bäst. Argumentera för varför de presterade bäst.</p> <p>För klassificering skall ni använda "WA_Fn-UseC_-Telco-Customer-Churn.csv". Denna innehåller observationer som intäkter från kunden, antal produkter, och andra data om kunden.</p> <p>Man vill hitta kunder som riskerar att falla bort. Detta representeras av den sista kolumnen, kallad "Churn", som kan ha värdena "Yes" eller "No". Bygg modeller som kan räkna ut Churn från de andra värdena.</p>
------------------	---

	Använd dessa observationer för att jämföra olika Machine Learning modeller, minst 2.
Hur löser ni uppgiften?	<p>Använd Numpy, Pandas och någonting som kan hantera en Notebook (ipynb) fil för att programmera i python, och använd modellerna i scikit learn för att göra undersökningarna.</p> <p>Du ska presentera en lösning med de modeller du vill använda samt argument om varför dessa modeller, med de parametrar du har valt, har gett bästa resultat.</p> <p>Förbereda data; t.ex. fixa inkorrekta eller ofullständiga värden, omvandla kategorisk data och gör de transformationer som behövs för användning i en modell.</p> <p>Dataset för första delen kan hittas på internet. UCI repository - https://archive.ics.uci.edu/ml/index.php Public datasets - https://github.com/awesomedata/awesome-public-datasets Google dataset search - https://datasetsearch.research.google.com/ Kaggle datasets - https://www.kaggle.com/datasets</p>
Struktur för arbetet	Uppgiften är individuell.

INLÄMNING OCH REDOVISNING

Inlämning	Projektet skall vara klart och färdigställt för bedömning senast den 16/6 . Skall lämnas som en .ipynb fil genom LearnPoint.
Redovisning	Arbetet skall inte redovisas.

BEDÖMNING OCH ÅTERKOPPLING

Bedömning sker med följande betygskriterier	<p>För godkänt (G) på projektarbetet skall följande krav uppfyllas:</p> <p>Skriv en förklaring av vad för typ av AI är bäst lämpad för denna data och varför.</p> <p>Gör en undersökning och jämförelse av minst två fungerande modeller per dataset.</p> <p>Diskutera även själva noggrannheten och beskriv valet av utvärdering av modellerna.</p> <p>Presentationen är tydligt uppdelad i sektioner med relevanta titlar.</p> <p>För väl godkänt (VG) på projektarbetet skall dessutom följande krav uppfyllas:</p> <p>Gör en undersökning, experimentering och jämförelse av minst fyra modeller per dataset. Variera olika parametrar och diskutera deras effekt på noggrannheten. (Experimentering.)</p>
---	--

Presentera bästa möjliga modell och argumentera varför den är bäst.
Förklara även vad resultaten innebär så att en icke teknisk åhörare
förstår.

Återkoppling

Återkoppling senast 2 veckor efter inlämningsdatum.

Lycka till!