**G Uppgift Uppgiften 1**

**Besvara följande teoretiska frågeställningar:**

**1.1. Vad är skillnaden mellan Regressionsproblem och Klassificieringsproblem?**

Regressionsproblem har en modell som är kontinuerlig kvantitet. Medan Klassificeringsproblem är en funktion i diskret klassetiketter. Klassificering är oorderad data medan regression är valda data. Beslutträd och logistisk regression är vissa typer av algoritmerna som Klassificieringsproblem kan användas utav. I Regressionsproblem används regressionsträdet och linjär regression. För mer klarhet så är regression predikterar kontunerliga utfall till exempel siffror. Medan klassificerade predikterar ett kategoriskt utfall t. ex JA/NEJ svar.

**1.2. Vad är ”supervised learning”? Kan du ge några exempel på modeller som används och potentiella tillämpningsområden? Vad är RMSE? Vad är en ”Confusion Matrix”?**

Med Supervised learing kan man skapa en prediktionsmodell utav flera variablerna (X1..Xn) som predikterar (Y).Genom en algoritm kan man beräkna hur X relaterar till Y variabel som blir en prediktionsmodell och brukar indelas med klassifikation och regression. Vissa modeller som används är t. ex. Linjär regression, som predikar för kontinuerliga eller numeriska variabler som t. ex ålder och pris. Regressionsanalys är väldigt populärt och den mest vanligaste att förekomma, relationen mellan x och y måste vara linjär för att modellen ska funka. Den vanliga metoden som man kan tillämpa utav linjär regression är ”least-squares” som reducerar summan av avvikelser i data, man får en förbättrade data.

Medan Djupinlärning är en del av maskininlärning, som skapas i flera lager, den är ordnande i neuronnät. Det gör att data blir processade genom flera steg som till slut får en prediktion till exempel när datorn ska tolka en bild, blir det till ett rutnät med siffror som går igenom neuronnätet. Med det kan dator manipulera bilden att se hörn och linjer osv. som till slut predikar en output. Den kan även analysera tabeller med data. Med djupinlärning underlättar den för arbete man måste göra på olika metoder, som är komplicerad och tar tid. Med djupinlärning får man en autonomi oavsett data.

RMSE står för ”Root Mean Square Error” det är en standardavvikelse för residualerna. Med andra ord kan man beskriva RMSE som ett mått på prediktion fel utifrån hur långt datapunkterna är från regressionslinjen. Den används oftast inom regressionsanalys för att fastställa experimentet resultat.

Confusion matrix är sammanfattningen för ett klassificieringsproblem, den delas upp efter varje klass med antalet korrekta och felaktiga antal. Den visar hur förvirrad klassificeringsmodellen är när den används. Den ger insikt vilka fel som gör och vilka fel typer. Med den metoden kan man får en noggrann klassificering.