

# Identifiering av Fake News med Hjälp Av Maskininläring Och Logistisk Regression

Belis Sabanovic  
Malmö Universitet  
2021-01-23

*Självständig fördjupning inom  
informations- och datavetenskap*

## Introduktion

På sistone har *falska nyheter* ökat markant. Den explosiva tillväxten av fejknyheter har visat sig ha flertalet negativa påföljder; bland annat ses dessa nyheter generellt som ett direkt hot mot demokratin [1]. Avsikten är oftast att vilseleda läsaren till att tro på något som inte är sant. Spridning av vilseledande information har också gjort det svårt att identifiera pålitliga nyhetskällor. Detta ökar behovet av verktyg som kan ge kännedom om tillförlitligheten hos innehåll, speciellt onlineinnehåll som bland annat förekommer i sociala medier. I nuläget finns diverse metoder för att någorlunda bromsa tillväxten av fejknyheter. Till exempel kan *maskininlärning* användas tillsammans med diverse maskininlärningsalgoritmer [2]. I denna uppsats kommer maskininlärningstekniker att användas tillsammans med algoritmen *logistisk regression* för automatisk identifiering av falskt innehåll på internet.

**Nyckelord:** Fake News/Fejknyheter, Maskininlärning, Logistisk Regression.

## Problemformulering

För att identifiera fejknyheter och avgöra tillförlitligheten hos onlineinnehåll kommer en logistisk regressionsmodell att skapas med hjälp av ett dataset. Modellens noggrannhet i procent kommer att tas fram för att skapa en uppfattning om dess tillförlitlighet mot det innehåll som användare i efterhand kommer att mata in. Därav har nedanstående fråga formulerats.

*Hur väl presterar algoritmen Logistisk Regression för att identifiera Fake News? Kan algoritmens prestation/resultat förbättras?*

Processen för att skapa en logistisk regressionsmodell ska således kunna svara på om en sådan modell/algoritm är tillräckligt bra för att avgöra en nyhets pålitlighet. Projektet ska också ta reda på om algoritmens resultat kan förbättras.

## Relaterat arbete

För att identifiera fejknyheter och avgöra tillförlitligheten hos onlineinnehåll har följande fråga formulerats:

## Möjliga lösningar

För att identifiera fejknyheter och avgöra tillförlitligheten hos onlineinnehåll har följande fråga formulerats:

## Befintliga lösningar

För att identifiera fejknyheter och avgöra tillförlitligheten hos onlineinnehåll har följande fråga formulerats:

## Slutgiltig lösning

För att identifiera fejknyheter och avgöra tillförlitligheten hos onlineinnehåll har följande fråga formulerats:

## Illustration

För att identifiera fejknyheter och avgöra tillförlitligheten hos onlineinnehåll har följande fråga formulerats:

## Reflektion

För att identifiera fejknyheter och avgöra tillförlitligheten hos onlineinnehåll har följande fråga formulerats:

## Bibliografi

För att identifiera fejknyheter och avgöra tillförlitligheten hos onlineinnehåll har följande fråga formulerats:

## Bilagor

För att identifiera fejknyheter och avgöra tillförlitligheten hos onlineinnehåll har följande fråga formulerats:

## Referenser

- [1] X. Zhou, R. Zafarani, K. Shu och H. Liu, "Fake News: Fundamental Theories, Detection Strategies and Challenges", *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM '19*, p. 836-837, Januari 2019. [Online]. Tillgänglig: ACM, <https://dl.acm.org/>. [Accessed: Feb. 23, 2021].
- [2] F. Nada, B.F. Khan, A. Maryam, N. Zuha och Z. Ahmed, "Fake News Detection Using Logistic Regression", vol 6(5), p. 5577-5579, Maj 2019. [Online]. Tillgänglig: Irjet, <https://www.irjet.net/>. [Accessed: Feb. 24, 2021].