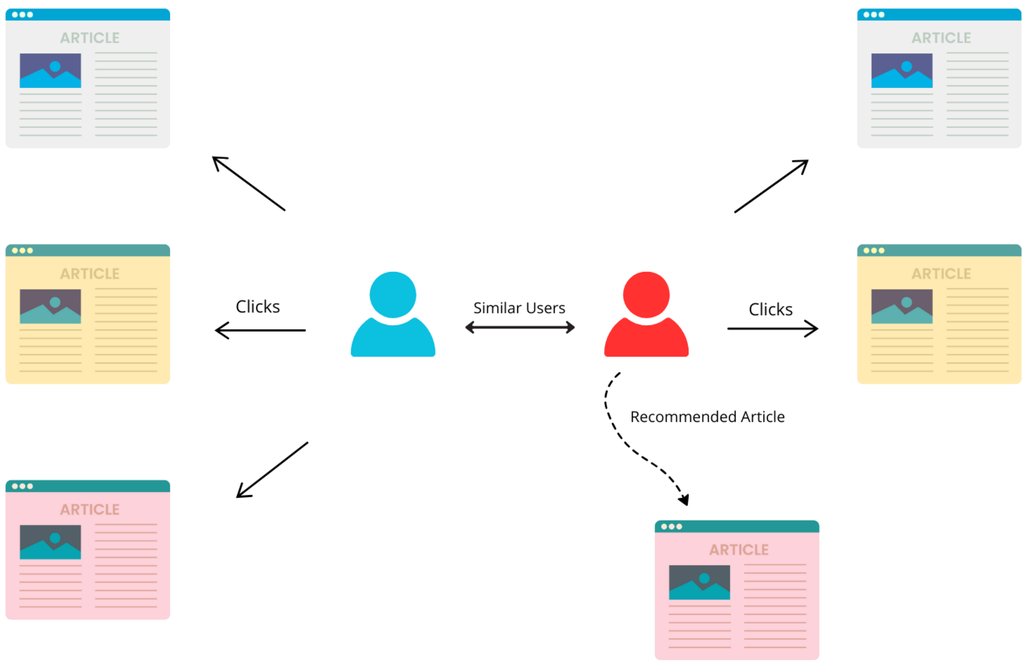
Personalized News Recommendation in the Sports Domain

Tord Berget Monclair, Master´s in Information Science, University of Bergen
tord.monclair@student.uib.no

Supervisor: Mehdi Elahi, University of Bergen
Co-supervisor: Thomas Husken, Bergens Tidende

Media Futures



Introduction

In today's digital landscape, personalized content recommendations play a crucial role in boosting reader engagement. This project focuses on implementing a recommendation system, utilizing state of the art techniques, tailored for local sport-related news articles on an online news platform. By analyzing user behavior, such as engagement and time spent on articles, this project aims to enhance user experience by recommending relevant content for each user. Leveraging collaborative and content-based filtering techniques, the goal is to predict and suggests articles that align with individual interests, thus supporting the news platforms mission to provide engaging, local sports content to its readers.

This project is a collaboration between the University of Bergen, MediaFutures and their industry partner Bergens Tidende. Data is provided by Bergens Tidende.

Research Questions

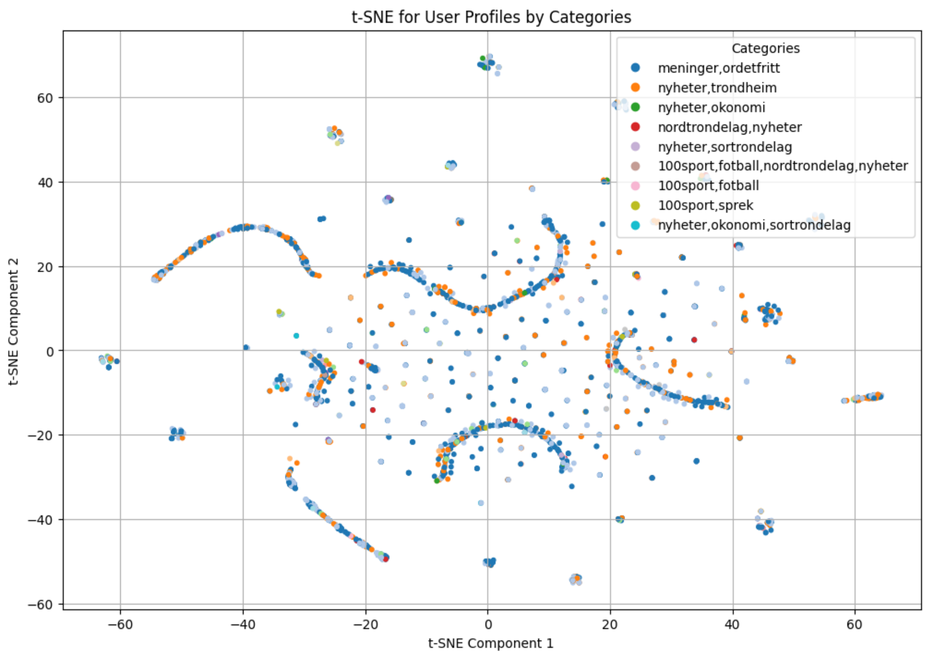
1. How can we develop a news recommendation approach that can learn from user data (in the sports domain) and generate appropriate recommendations for users?
2. How does the performance of the developed news recommendation approach compare to a traditional technique?

Method

To familiarize myself with the currently available tools, the first part of the project has been to build a recommender system based on the Adressa dataset from NTNU. This dataset was chosen as it contains some similarities with the dataset provided by BT and could act as a good starting point to conduct some offline evaluation and exploration. At this point in time, I have conducted an offline evaluation of matrix factorization-based models which predicts the expected reading time for a given user-item pair. The results illustrated in the figure below reflect improvement over the considered baseline:

Model	Precision@5	Recall@5
Random	0.305074	0.294722
SVD	0.528155	0.566533
SVDpp	0.528117	0.566443

The next part of the thesis will focus on improving the recommender and tuning it to work on the BT dataset, which contains more rich data on both users and articles. By incorporating the content of articles in the recommendation, the goal is to ultimately improve the current personalization approach for local sports-related articles.Several suitable tools and Python libraries will be applied to test whether more complex models provide a potential improvement over traditional techniques. Illustrated below is a figure showing users clustered by their geographic location and device type for the Adressa dataset:



This thesis aims to potentially conduct an A/B test in collaboration with BT to test whether or not the new approach improves engagement over a baseline approach.

PARTNERS

amedia Schibsted 2 NRK Universitas Osloensis MDCCCXI Bergens Tidende nb.no VIZRT WOLFTECH NORCE UiS Faktisk.

HOST

UNIVERSITAS BERGENSIS UNIVERSITY OF BERGEN

FUNDED BY

This research is funded by SFI MediaFutures partners and the Research Council of Norway (grant number 309339).

Forskningsrådet Norwegian Centre for Research-based Innovation