

## BSc Project – Preliminary Problem Statement

My bachelor project in Data Science will assess and compare different recommender systems for podcasts. In particular, podcasts published by the Danish Broadcasting Corporation (DR) in their own audio platform DR LYD. I will focus on recommender systems based on explicit feedback and textual content, respectively.

Users are listening to podcasts in various ways, both in terms of frequency and purpose. This makes the task of recommending new podcast episodes for each user a challenging task. Each recommendation method might capture different aspects of the complex recommendation task. Thus, it could be interesting to compare the content-based system, and the one based on explicit feedback, e.g. collaborative filtering, with some baseline recommendation technique. Such baseline could be recommending the most popular episode, alternatively within a specific genre or group of users.

From the comparison described above, it could be interesting to compare the performance, in terms of accuracy, of a content-based system with collaborative filtering. For the content-based system, the level of detail of the content can be varied. This can be done by first only including the episode title, then the episode description and finally a full transcription of the episode. A content-based recommender system might also be helpful in cold-start scenarios, where the amount of explicit user feedback is limited.

To summarise, I'm considering the following research questions, where performance refers to accuracy of the system:

- *How does a content-based recommender system perform compared to collaborative filtering for podcast episode recommendations?*
- *How does the granularity of the content affect the performance of a content-based recommendation system?*
- *How does a content-based recommendation system perform for new users in a cold-start scenario?*

It could also be worth discussing the purpose and prevalence of recommender systems in a public service setting. Should the recommendations be highly personalized, or is directing the user towards unexplored topics and genres a higher priority? This could be a topic for the discussion section.