KNIME Student Challenge: Gathering Insights from Movie Reviews with Text Processing and AI

Due: December 3, 2025, by 5:00 pm
Total Points: 100

Objective

In this project, **teams** will analyze a dataset of movie reviews (available here) to determine their sentiment and to identify groups of reviewers with similar reactions and tastes for movies.

This assignment develops your ability to communicate data insights with empathy and precision by tailoring messages to audience needs, rather than using one-size-fits-all communication.

Task

Students should experiment with traditional text processing and machine learning techniques and compare them with generative AI techniques. **How do these techniques compare?** What insights can be uncovered from the data, in terms of sentiment and taste segments? Students are encouraged to also experiment with data visualization techniques and dashboards to better communicate their findings. Solutions should be implemented in KNIME Analytics Platform.

Deliverables – recording, report, and KNIME workflow

A recording of 5-7 minutes (7 minutes max, no exceptions) that includes:

1. Overview of the solution with its key decisions
2. Most relevant parts of the workflows

A report (5-page max, no exceptions) that includes:

1. **Techniques used** – What text processing, machine learning, and generative AI techniques were used, and how do they compare?
2. **Insights –** What are your findings with respect to users' tastes and how they are segmented? What are the top movies and genres in the dataset? How about the bottom movies and genres? What do they have in common? Are there traits in movies that correlate with worse or better reviews?

A KNIME workflow, or set of workflows, containing:

1. Data exploration.
2. Preprocessing of textual data. Use the KNIME Textprocessing extension.
3. Application of traditional sentiment analysis techniques to predict the overall sentiment of reviews. Examples can be found here.
4. Application of generative AI and prompting techniques to predict the overall sentiment of reviews. The exploration of aspect-based sentiment analysis is encouraged here as well (see example here). Feel free to use the KNIME AI Extension.

5. Application of traditional machine learning and generative AI techniques to identify segments of reviewers with similar taste for movies.
6. Creation of a data app with visualizations that convey important insights from your analysis.

Other things to keep in mind before submitting

1. All files should be named as Team##-name of the file. E.g., Team01-Recording, Team01-Report, Team01-KNIME Workflow
2. Make sure all three files are included in the same attempt.
3. Make sure you submit before the deadline – 12/3/2025 by 5:00 pm CT.
4. If you need to share an API key, make sure that it is submitted as well. I will not reach out if the file does not run.
5. EXTRA – you can submit a PPT file that you use when presenting/recording as a separate file. Make sure you name it correctly.
6. All file names should start with the team number.
7. All submissions should have the team number and members names.
8. Peer review should be submitted separately by individual team members.