

Arbeitsjournal Reto Lämmli

Name Reto Lämmli

Datum	Stunden	Tätigkeit
Saturday, December 13, 2025	1.5	Experimented with a simple regression model using only my own LinkedIn data and Gemini.
Monday, December 15, 2025	3.0	Research new LLM-based LinkedIn algorithm (https://arxiv.org/html/2510.14223v1) Setup LinkedIn Post for Data Collection
Tuesday, December 16, 2025	1.5	Moderating LinkedIn Respondents, Sharing Instructions (https://forms.gle/ye5LNDvww3BWTTS56)
Wednesday, December 17, 2025	1.0	Moderating LinkedIn Respondents
Saturday, December 27, 2025	5.0	Rewriting of Project Description, to include Semantic Alignment of Profile vs. Post Content Sorting and enriching data provided by LinkedIn users using LinkedIn Post Extractor Experimenting with ChatGPT agent to fetch Reposts.
Sunday, December 28, 2025	5.0	Tried and Failed to execute "Fetch Reposts" with ChatGPT Agent and Manus.ai to extend all Shares.csv with Repost number. It just doesn't scale. Continue only with Likes and Comments. Write Python script to: - extend Shares.csv with Network Size at posting time using Connections.csv - extend Shares.csv with Author profile using Profile.csv - consolidate all Shares.csv into "Master Shares" file - calculate Semantic Alignment between Posts and Author profile using all-MiniLM-L6-v2 vs. paraphrase-multilingual-MiniLM-L12-v2
Monday, December 29, 2025	2.0	Review and optimized scripts
Tuesday, December 30, 2025	3.0	Introduced new features: - Post Content Length - Hour of day - Day of Week Changed target variable from Relative Engagement prediction to a binary classifier (Is_High_Performing) Results: y = Relative Engagement stable around 0.2 - 0.3 y = Is_High_Performing stable around 0.65
Wednesday, December 31, 2025	3.0	Experimented with split of dataset before vs after LinkedIn algorithm change in 2025. In 2025, Semantic_Alignment gains in influence and Post_Content_Length loses its influence. The split test shows the most significant change of features after August 2025.
Saturday, January 03, 2026	5.0	Doubled the Headline, to increase the significance of the profession in the semantic alignment. Tried new feature "Posts last 7 days". No significant impact. Cleaned and optimized the scripts. Researching more feature engineering: hook length, emoji count, hash tag count, sentiment, external links, etc. Reminded of people, to send data. Currently at 5k posts from 30 users Currently at Accuracy of 66%
Sunday, January 04, 2026	3.0	Implement and tested new features : features = ['Network Size', 'Semantic Alignment', 'Sentiment Score', 'Hook Length', 'Emoji Count', 'Hashtag Count', 'Linebreak Count', 'Mention Count', 'Posts Last 7D', 'Post Content Length', 'Hour', 'Day Of Week', 'Has Image', 'Has Video'] Also tested XGBoost. Plateaus at 67%. Randomforest reaches 69% accuracy. F1-recall for High Performing is 74% though.
Monday, January 05, 2026	2.0	Imported more new user data Hyperparameter finetuning for XGBoost => No major effect Continued to feature engineer and finetune More or less same quality. No new major breakthrough. Excluded Mention Count and Posts Last 7D again.
Tuesday, January 06, 2026	1.0	Changed Sentiment Analysis Model to nlptown/bert-base-multilingual-uncased-sentiment to have multilingual support Quality is good, but effect in Model minimal
Wednesday, January 07, 2026	1.0	Importing additional data Finetuning model with turning on/off features Remove "Posts Last 7D" as a feature because due to the data cleanup, this is not a consistence feature
Thursday, January 08, 2026	1.0	Kick-off structure for documentation
Friday, January 09, 2026	3.0	Experimented with 3 classes: Low - Medium - High Performance. The middle class is unstable, therefore reverted back to binary class. Built a SHAP Graph to understand how the model features influence the prediction score. Started to structure the documentation using NotebookLM
Saturday, January 10, 2026	2.0	Introduced another feature: Link count Implemented SHAP analysis
Sunday, January 11, 2026	6.0	Separated python scripts to be more modular. Less regenerating of semantic and sentiment analysis Introduced a virtual environment Visualized SHAP Recreated comparison of SHAP of All vs. until March vs. from March
Monday, January 12, 2026	1.0	Imported new data arrival and checked the model again.
Tuesday, January 13, 2026	2.0	Reflecting on data leak, because the train / test split mixes posts from all users. Explored kfold but data amount is limited to successfully use it. Discussed with Tobias Merinat the issue. He said he doesn't believe there is a data leak since no personal features like user id is used.
Sunday, January 18, 2026	3.0	Writing documentation. Defined Table of contents. Wrote Abstract, Introduction and Literature Review.
Monday, January 19, 2026	1.5	Wrote Chapter Data Acquisition and Data Preprocessing
Thursday, January 22, 2026	1.0	Wrote Chapters Dataset Generation, Model Selection and Model Training
Sunday, January 25, 2026	3.0	Wrote Chapters Results and Model Evaluation, Concept Drift, Summary of Findings
Monday, January 26, 2026	1.5	Document finalization, creating figures like Learning curve, dataset distribution etc.
Tuesday, January 27, 2026	2.0	Challenged Learning curve and realized that strong overfitting is happening. Started to retune hyperparameters.
Wednesday, January 28, 2026	4.0	Went back to finetune hyperparameters to achieve a better Learning curve since previous one turned out to be overfitting. Rewrote every chapter with the latest findings and results
Thursday, January 29, 2026	3	Created official Citations, List of Figures and List Tables, Started Review Process
Friday, January 30, 2026	4	Full review and refinement of final draft
Saturday, January 31, 2026		Proof-reading.
Sunday, February 01, 2026	2	Extended Random Forest as baseline into the documentation
Sunday, February 01, 2026	3	Finalizing proof-reading, signing off documentation and upload on Iliia Cleaned up Github, separated Data Folders into a private Repo.
Monday, February 02, 2026		Kicked off presentation, using NotebookLM as a baseline. Started to outline a draft.
Tuesday, February 03, 2026	6	Finalized presentation
Total	88.0	