

Research Project: Influence of demographic factors on Individuals news article preference

Christof Wiesinger, Michael Stark, Ruslan Shkira

Challenge

The aim of the project:

- Understand relevant news topics
- Help news outlets better understand their target audiences
- Help consumers understand their news preferences
- Predict the preferences of readers depending on the core factors



Project goals:

- Create a website on the data, provided by the users
- Train a supervised model for regression
- Visualise the data for website users

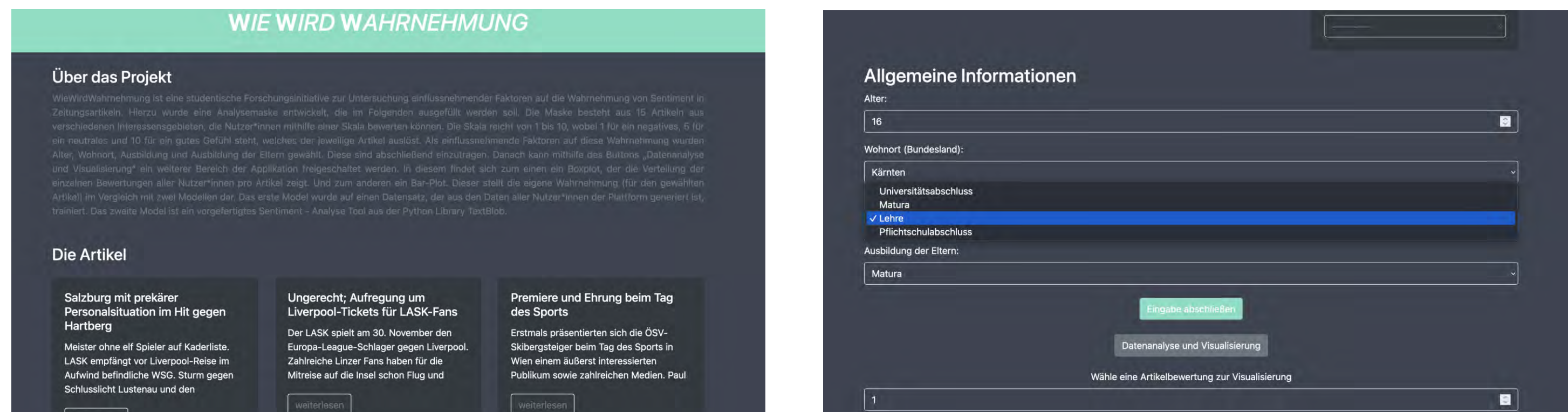
Dataset

- The test data has been **created from different research papers**, investigating reading preferences of different people
- **50 instances** for participants of different age groups, from different locations and educational backgrounds
- **4 attributes:** Age, Location, Level Of Education and Parents' Education, followed by grading of newspaper articles

| id | rating1 | rating2 | rating3 | rating4 | rating5 | rating6 | rating7 | rating8 | rating9 | rating10 | rating11 | rating12 | rating13 | rating14 | rating15 | rating16 | age | location | education | parent_education |
|----|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|----------|----------|----------|----------|----------|----------|-----|----------|------------------------------------|------------------------------------|
| 1 | 7 | 4 | 9 | 10 | 3 | 1 | 7 | 2 | 8 | 10 | 7 | 4 | 1 | 10 | 6 | 4 | 64 | Wien | Bachelors/Masters/Doctorate degree | Bachelors/Masters/Doctorate degree |
| 2 | 5 | 8 | 10 | 8 | 5 | 2 | 6 | 4 | 9 | 9 | 6 | 2 | 2 | 9 | 5 | 5 | 53 | Wien | College | Master's degree |
| 3 | 3 | 6 | 10 | 10 | 5 | 2 | 5 | 2 | 7 | 10 | 5 | 5 | 3 | 8 | 3 | 7 | 70 | Wien | High school degree | No education |

Data Understanding & Preparation

- Study existing papers on article preferences and focus on main attributes
- Choose 5 main topics and 3 subtopics of articles
- Create boxes for participants' input
- Analyze initial data, provided by the participants

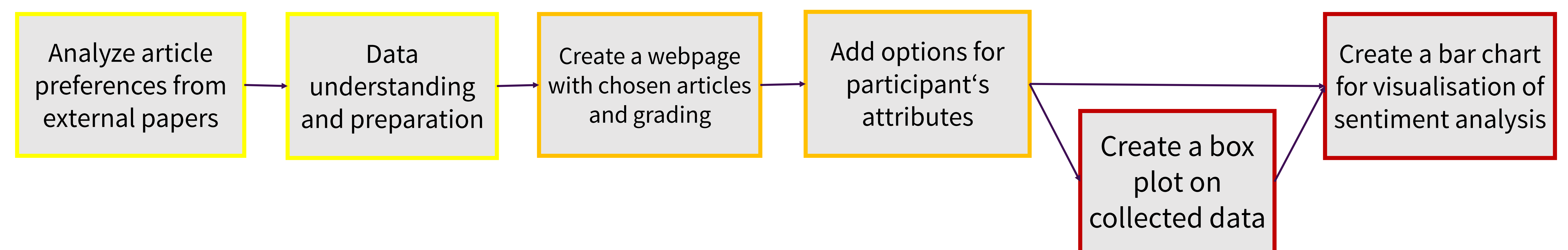


Visualisation

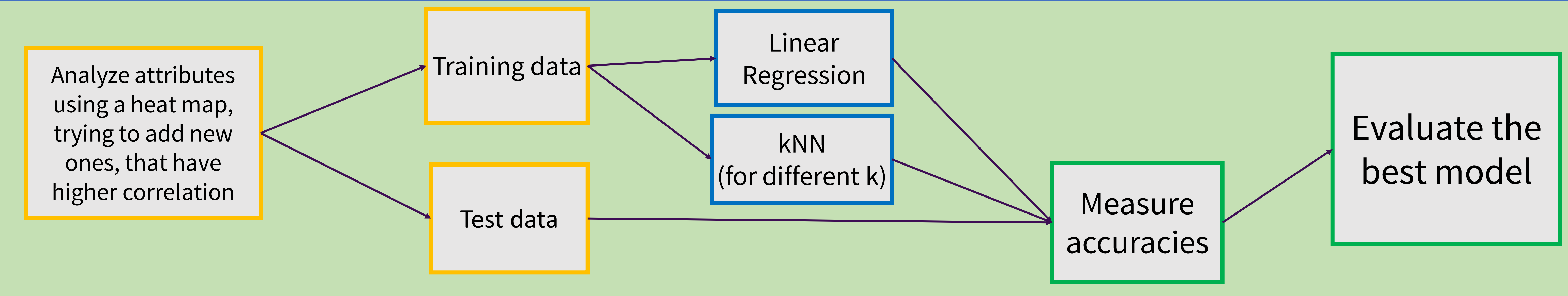
- **Box plot** to showcase article preferences of all participants
- Additional **bar chart** to evaluate how different perceptions are
- Future additional models and visualisations:
 - **Linear Regression**
 - **k-Nearest Neighbour Regression** for different k



Workflow

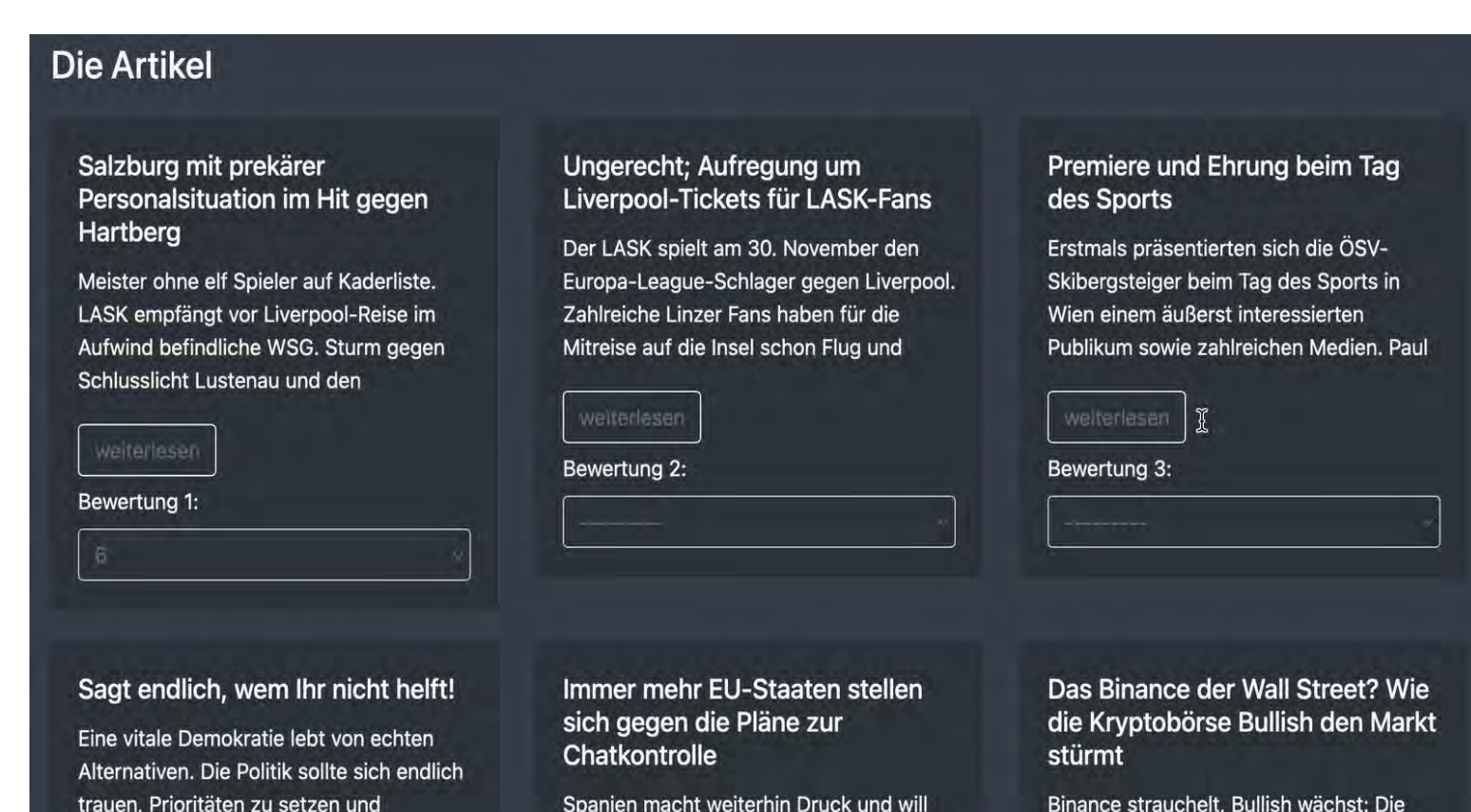


Future roadmap of the project



Results so far

- **Functioning website** for research purposes
- **Auto-updating data analysis and dataset**, depending on participants' input
- We also considered how perception would change if we chose a different wording of the headline of the same article



Conclusion

- Website prepared for data gathering and initial data analysis with visualisation and possibility to expand the research
- There is still a lot to contribute to the project with sufficient financial and resource investment
- Additional attributes and additional models can prove considerably useful for many news outlets

