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

<table>
<thead>
<tr>
<th>Age</th>
<th>Location</th>
<th>education</th>
<th>parent education</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>College</td>
<td>Bachelor’s</td>
<td>Bachelor’s</td>
</tr>
<tr>
<td>35</td>
<td>School</td>
<td>Bachelor’s</td>
<td>Master’s</td>
</tr>
<tr>
<td>45</td>
<td>University</td>
<td>Bachelor’s</td>
<td>Bachelor’s</td>
</tr>
<tr>
<td>55</td>
<td>Work</td>
<td>Bachelor’s</td>
<td>Master’s</td>
</tr>
</tbody>
</table>

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

Create a webpage with chosen articles and grading
Add options for participant’s attributes
Create a bar chart for visualisation of sentiment analysis
Create a box plot on collected data

Future roadmap of the project

Analyze attributes using a heat map, trying to add new ones, that have higher correlation

Training data
Linear Regression
kNN (for different k)
Test data
Measure accuracies
Evaluate the best model

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