REI603M - Assignment 3

Your feedback shapes the course! To help continually improve and tailor future assignments to your needs, I invite you to participate in a brief survey, accessible here (URL: `https://forms.gle/g2txWcSisnu8CSE1A`).

Submission Instructions: Please submit this assignment via Canvas. For multiple files, kindly compress them into a single `.zip` file. Please name your submission file as

`REI603M_Assignment3_YourName_YourPartnerName.zip`

Data Preparation and Feature Engineering

In the past week, we explored data preparation and feature engineering. It's essential to not only be familiar with these techniques but also to apply them effectively. This assignment is an opportunity for you to demonstrate your understanding and skills in this area.

Your task is to present a comprehensive overview of the methods you applied to your chosen dataset. This is also a great time to expand your dataset if it's currently limited in size.

To guide your exploration, here are some valuable resources:

- Explore imbalanced-learn, a Python package for handling imbalanced datasets. A useful starting point is this Kaggle kernel.
- Investigate various feature selection techniques. We discussed Boruta in our lecture; consider combining it with tools like Featuretools or feature-crossing strategies.
- For feature augmentation/synthesis, particularly for images, refer to this Kaggle kernel.
- Explore mixup as a way to combine existing examples. Here is an implementation in Keras, but you might need to perform this implementation yourself.
- Dimensionality reduction: Explore linear matrix factorization methods like PCA or NMF, and consider using UMAP for structural analysis of your dataset.
- Featuretools, an open-source Python framework for automated feature engineering. Numerous example notebooks are available for guidance.
- Utilize GPT-4 for personalized suggestions on feature engineering tailored to your project and apply the suggestions to your problem.

Focus is key, so choose methods relevant to your dataset and project scope.

The Task

Data Experiment: Apply the suggested data preparation and feature engineering techniques to your dataset.

Model Testing: For those working with labeled datasets, assess the impact of your feature engineering by training models on both the original and modified datasets. Ensure to use a test set for evaluation.

Data Improvements: If you lack labeled data or need more data, focus on labeling or data generation in addition to the feature engineering efforts. If you go for this route, and you cannot train a model, then you should at least perform some qualitative evaluation of the feature engineering you pursued.

Prompt Engineering: For projects involving Language Models like GPT-4, feature engineering may not be as relelvant (although it might still be). In case you cannot find a good way to apply standard feature engineering techniques, experiment with prompt engineering to enhance output quality. Select 2-3 prompt variations for detailed evaluation.

Reflection After doing your experiment, introspectively assess your work, acknowledging strengths, limitations, and potential areas for improvement or further research.

Submission Requirements

Your submission should include:

- The script (notebook or `.py` files) used for your analysis.
- Presentation slides.

Your presentation should be succinct, clear, and within 5-7 minutes. Slides are recommended for ease of transition between presenters.

Peer Evaluation

Participate in the peer evaluation process using the following rubric. Please refrain from evaluating yourself or your group members.

- **Application of Techniques (0-5 points):**
  - **0 points:** No application of techniques or irrelevant to the project's scope.
  - **1-2 points:** Minimal application of techniques with limited effectiveness or relevance.
  - **3-4 points:** Effective application of techniques, showing good integration with the project's objectives and data characteristics.
  - **5 points:** Outstanding application of techniques, demonstrating a high level of skill and a significant positive impact on the project.
- **Analysis and Results Interpretation (0-5 points):**
  - **0 points:** No or incorrect analysis and interpretation of results.
  - **1-2 points:** Basic analysis with superficial or partially correct interpretation of results.
  - **3-4 points:** Well-conducted analysis with clear and correct interpretation of results, showing relevance to the project's goals.

  - **5 points:** Exceptional analysis and insightful interpretation of results, providing deep understanding and potential future implications.

- **Innovation and Creativity (0-5 points):**
  - **0 points:** No evidence of innovation or creativity.
  - **1-2 points:** Minimal innovation or creativity in approach or execution.
  - **3-4 points:** Good demonstration of innovative and creative thinking in problem-solving and project execution.
  - **5 points:** Outstanding innovation and creativity, significantly enhancing the project's value and approach.
- **Critical Evaluation and Reflective Analysis (0-5 points):**
  - **0 points:** No critical evaluation or reflection on the process and outcomes.
  - **1-2 points:** Limited critical evaluation, with superficial reflection lacking depth or insight.
  - **3-4 points:** Good critical evaluation and reflective analysis, showing understanding of the project's strengths and limitations, and identification of areas for improvement or further exploration.
  - **5 points:** Exceptional critical evaluation and reflective analysis, demonstrating deep understanding and insight into the project's process and outcomes, including thoughtful consideration of future directions and potential impact.
- **Presentation and Communication (0-5 points):**
  - **0 points:** Incoherent or poorly constructed presentation.
  - **1-2 points:** Basic presentation, lacking in clarity or engagement.
  - **3-4 points:** Good presentation, clear and well-organized, with effective communication of key concepts.
  - **5 points:** Exceptional presentation, highly engaging and informative, excellently communicating complex ideas in an accessible manner.

Deadline Reminder: Your assignments must be submitted by 18:00 on February 7th.

We wish you success with this assignment! For any inquiries, reach out via Ed or email `hafsteinne@hi.is`.