Annex G. Justification for the taken decisions in the methodology

**Table 3.3.** Justification for the taken decisions in the methodology

|  |  |
| --- | --- |
| **Decision regarding** | **Justification** |
| Imbalanced data set | * Based on literature: [Winata et. al. (2015)](#Winat) ; [Charte et. al. (2013)](#Charte); [Liu et. al. (2015](#liu15)); * Sub-chapter 2.1.7. – application of One Vs One approach (experiment); * Sub-chapter 2.1.9. – application of *balanced weights method*(experiment); application of the stratification(experiment); * Sub-chapter 2.1.11. – application of tree-based methods with boosting(experiment); * Sub-chapter 1.2.1. – comparative studies, experiments made by other researchers; |
| Correlated features in modeling | * Based on literature: [Barrera (2021](#Barrera)); [De Luca (2020)](#Luca); [Cook (1995](#Cook)); [Molnar (2022)](#Molnar); * Sub-chapter 2.1.9. – application of NN; application of the stratification(experiments); * [Goštautaitė (2019b)](#PCA19); * Sub-chapter 1.2.1. |
| Correlated features and SHAP | * Based on literature: [Pedregosa et. al*.* (2011)](#Pedregosa); [Mase et. al*.* (2019)](#Mase); [Aas et. al. (2021);](#aas) * Subs-chapter 1.11.3. – application of the SHAP method; * Sub-chapter 2.1.13. – experiment; |
| Correlation between labels | * Based on literature: [Chiang et. al*.* (2012)](#Chiang); [Dery et. al*.* (2021)](#Dery); * Sub-chapters 1.2.3.  application of label ranking; * Sub-chapter 2.1.4., 2.1.5.  application of the Label Powerset and classifier chains for the correlated labels(experiment); * Sub-chapter 1.2.3.  KNN with label ranking (experiments ([Chiang et. al., 2012](#Chiang))); * Sub-chapter 2.1.11.  KNN with label ranking (experiments ([Cause, 2023](#CAUSE))); * [Table 1.6.](#tab6) Categorization of the multi-label classification algorithms based on the degree of correlations among labels; * Sub-chapter 1.2. – comparative studies, experiments made by other researchers; * Chapter “Problem formulation”- classification approach for scenarios with correlated labels; * Sub-chapter 2.1.6. – classification approach for scenarios with correlated labels; |
| Missing values in a data set | * Sub-chapter 1.2.3. – training with missing values; * Sub-chapter 2.1.9. – *XGBoost* can handle missing values, but additional methods (imputation methods) have to be applied in the case of NN (experiments); |
| Sparse data in a data set | * Sub-chapter 1.2.1. – comparative studies, experiments made by other researchers; * Sub-chapter 2.1.7. – and for handling sparse data (experiment); * Sub-chapter 2.1.9. – comparison of the experimental results (NN); |
| Sparse labels | * Sub-chapter 2.1.9. – poor NN performance since the labels are distributed sparsely (experiments); |
| Use of the learning style model developed in the cognitive theories | * Sub-chapter 3.1.  clustering vs classification; * Sub-chapter 2.1. – application of supervised learning (experiments); |
| Supervised vs unsupervised | * Based on literature: [Pedregosa et al. (2011)](#Pedregosa); [Kim (2015)](#Kim15); [Nakashe (2018)](#Nakashe); * Sub-chapter 1.2.2.-1.2.3., 2.1.3.-2.1.10.; * Sub-chapter 1.2.2.-1.2.3., 2.1.3.-2.1.9. – supervised classification methods; * Sub-chapter 1.2.5. – unsupervised machine learning modeling; * Sub-chapter 2.1 – application of supervised learning methods (experiments); * Sub-chapter 2.1.10.  KNN with label ranking (experiments [(Cause, 2023)](#CAUSE)); * Based on the experimentation conducted by [Nakashe (2018)](#Nakashe) – K-Means application (unsupervised); |
| Generative vs Discriminative | * Based on literature: [Hewitt (2018)](#Hewitt); [Rufai (2020)](#Rufai); [Griffiths (2019)](#Griffiths); [Bishop *et. al.* (2007)](#Bishop); [Bach *et. al.* (2017](#Bach)); [Kim (2015](#Kim15)); [Generative models (2016)](#Karpathy); * Sub-chapter 1.2.2. – generative vs discriminative; * Sub-chapter 2.1.12. – application of the generative approach, Gibbs sampling; * Sub-chapter 3.1. – application of generative and discriminative approaches; * experiments conducted; |
| Prediction of a single unique label combination vs predictions for each label (learning style dimension) | * Sub-chapter 2.1.7. – application of the “One vs One” strategy (experiment); * Sub-chapter 2.1.4. – application of the Label Powerset method; (experiment); |
| Inherently interpretable model vs model agnostic interpretability of the model | * Based on literature: [Goštautaitė et. al. (2022)](#MLDG22); [Mathworks (2022)](#Mathworks), [Kim (2015)](#Kim15); [Billiau (2021)](#Billiau); [Molnar (2022](#Molnar)); [Mase et. al. (2019)](#Mase); [Aas et. al. (2021).](#aas) * Sub-chapter 1.11.3. – application of SHAP method; * Sub-chapter 2.1.13. – experiment; |
| Epistemic and aleatory uncertainty | * Based on literature: [Indrayan (2020](#Indrayan)); * Sub-chapter 1.2.5. |
| Conditional independence | * Based on literature: [Goštautaitė (2019)](#GBN19); [Deventer (2004)](#Deventer); * Sub-chapter 1.2.4. – generative supervised machine learning model; * Sub-chapter 2.1.12. |
| Existence of invariants/inductive bias | * Sub-chapter 2.1.9. – NN modeling (experiment); |
| Multi-label classification algorithms that do not require meta-estimators | * Sub-chapter 2.1.11. – application of Ridge classifier(experiment); |
| Size of the data | * Sub-chapter 2.1.9. – size of the data set and the impact on the performance of the model; * Sub-chapter “Neural networks” – the impact of the size of the data set on the generalization ability; * Sub-chapter 2.1.9. – application of NN (experiment); * Sub-chapter 2.1.10. – comparison of the experimental results (NN); |
| Data normalisation, scaling | * Sub-chapter 2.1.9. – NN performance with and without scaling (experiment); |
| Selection of hyperparameters of the model | * Sub-chapter 2.1.6. – application of *GridSearch* for the selection of hyperparameters(experiments); * Sub-chapter 2.1.9. – application of *GridSearch* method for selection of hyperparameters for NN (experiment); |
| Handling outliers | * Sub-chapter 2.1.2., *Annex T.docx* - use of statistical methods (e.g., Z-score, IQR) or domain knowledge to identify outliers; remove outliers if this won’t lead to information loss; * Sub-chapter 1.2.2. – in NN using activation functions that are less sensitive to outliers, such as the rectified linear unit (ReLU) or its variants; * Sub-chapter 2.1.9. – NN performance with and without scaling (experiment); |
| Role of the learning rate in neural network training | * Sub-chapter 2.1.9. – application of *GridSearch* method for selection of hyperparameters for NN (experiment); |
| *BR* vs *One vs Rest* | * Sub-chapter 2.1.8. |