VILNIUS GEDIMINAS
TECHNICAL UNIVERSITY
FACULTY OF FUNDAMENTAL SCIENCES

VIRTUALIŲ MOKYMOSI APLINKŲ PERSONALIZAVIMAS IR VERTINIMAS: KARKASAS, SKIRTAS BESIMOKANČIOJO MOKYMOSI STILIAUS MODELIAVIMUI IR MODELIO VERTINIMUI BEI MODELIO PROGNOZIŲ INTERPRETAVIMUI

PERSONALIZATION AND EVALUATION OF VIRTUAL LEARNING ENVIRONMENTS: FRAMEWORK FOR STUDENTS' LEARNING STYLE MODELLING, EVALUATION OF A MODEL AND INTERPRETATION OF MODEL PREDICTIONS

doktorantė **Daiva Goštautaitė**

Tikslai/Aims

- The thesis aims of guiding the modelling process and helping to systematically organize the information and ideas around the problems being solved in student's learning style modelling by developing a common reusable universal supportive structure around which student's learning style model may be created.
- Focusing on the learning style modelling journey, not at any single model, the framework provides a common foundation so that researches, data scientists and model developers don't have to redo it from scratch when modelling learning style of a learner.
- Addressing common modelling challenges or requirements, the framework provides structured approach to student's learning style modelling and enables to select and adjust methods for multidimensional student's learning style detection and interpretation under given conditions and restrictions.

Methodology -> Development of a framework

- **methodology** refers to the systematic approach, techniques, and procedures used to conduct research, solve problems, or achieve specific objectives. It outlines the steps and processes involved in carrying out a study or project.
- **framework** provides a conceptual structure or scaffolding for organizing, understanding, and addressing complex problems or phenomena. It outlines key concepts, principles, relationships, and components relevant to a particular domain or field.
- **development of a framework** involves applying scientific principles and methods to organize, explain, and advance knowledge within a particular domain, making it a scientific activity.
- **development of a framework** can indeed be considered a scientific activity, depending on the context and the approach taken. A framework typically involves a systematic approach to organizing and structuring concepts, methods, or processes within a particular field or discipline. This often involves research, experimentation, analysis, and validation of the framework's effectiveness. In many scientific fields, frameworks serve as essential tools for understanding complex phenomena, organizing data, guiding research, and making predictions.
- The scientific rigor involved in developing a framework often includes:
  - **Research:** Gathering existing knowledge, theories, and practices related to the topic.
  - **Conceptualization:** Formulating a theoretical foundation or conceptual framework to guide the development process.
  - **Design:** Creating the structure and components of the framework based on the conceptualization.
  - **Testing and Validation:** Evaluating the framework through empirical studies, simulations, or other methods.
  - **Refinement:** Iteratively refining the framework based on feedback and new evidence.

Uždaviniai/Tasks

- Research supervised and unsupervised machine learning methods for clustering and classification and identify those that are most suitable for students' learning style modelling;
- Investigate experimentally the application of various machine learning methods for learner's learning style modelling under certain conditions and restrictions, evaluate the skill and performance of the models developed and compare the results both with each other and with the results obtained by other authors;
- Select interpretation method(s) for supervised learning style models;
- Create and validate a framework for selection of methods appropriate for the prediction of learning style of a student (or a group of students) and interpretation of the prediction results.
- Propose ways to integrate learning style model(s) into the virtual learning environment;

Scientific approach

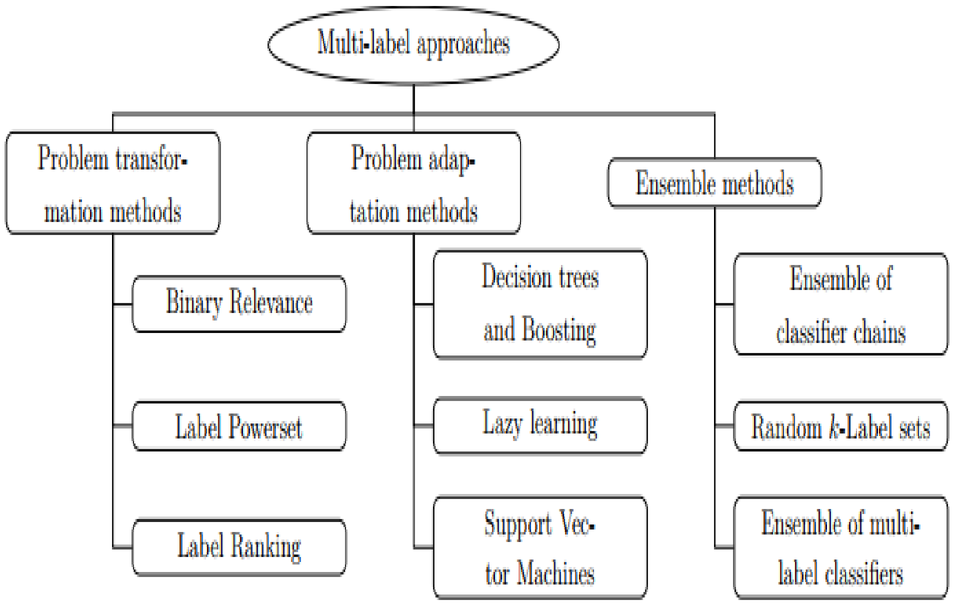
- **constructive research approach** (theoretical analysis, thinking + heuristic innovations, are characteristic of constructive research approach) for supervised classification. It leans on *pragmatism* (Lukka, 2000; Pasian 2015): considers words and thought as tools and instruments for prediction, problem solving and action, and rejects the idea that the function of thought is to describe, represent, or mirror reality; *rejects the idea of absolute truth or fixed principles and instead focuses on the usefulness and effectiveness of ideas in solving problems and guiding action*.The central notion of this approach, *the (novel) construction*, is an abstract notion with great, in fact, infinite number of potential realizations. All human artefacts, such as models, diagrams, plans, organisation structures, commercial products and information systems designs, are constructions. It is characteristic of them that they are invented and developed, not discovered. Novel constructions bring forth new reality. The core features of constructive research aquired that it (Lukka 2003):
  - focuses on real-worl problems felt relevant to be solved in practice;
  - produces an innovative construction meant to solve the initial real worl problem;
  - includes an attempt for implementing the developed construction and thereby a test for it's practical applicability;
  - implies a very close involement and co-operation between the researcher and practitioners in a team-like mannear in which an expierential learning is expected to take place;
  - is explicitly linked to a theorethical prior knowledge;
  - pays particular attention to reflecting the empirical findings back to theory.
- **constructive type of approach** used in the dissertation demands a form of validation that *doesn't need to be quite as empirically based as in other types of research like exploratory research*. Nevertheless *the conclusions have to be objectively argued and defined. This may involve evaluating the "construct" being developed analytically against some predefined criteria or performing some benchmark tests with the prototype*. The term "construct" is often used in this context to refer to the new contribution being developed.
- **conceptual resaearch** is descriptive and theorethical, it produces new knowledge primarily through the 'method of reasoning' (Lukka 2003). This approach was applied for the part of reseach that concers the generative statistical classification approach, unsupervised learning, exemplar-based methods and integration of the learner model into virtual learning environment.

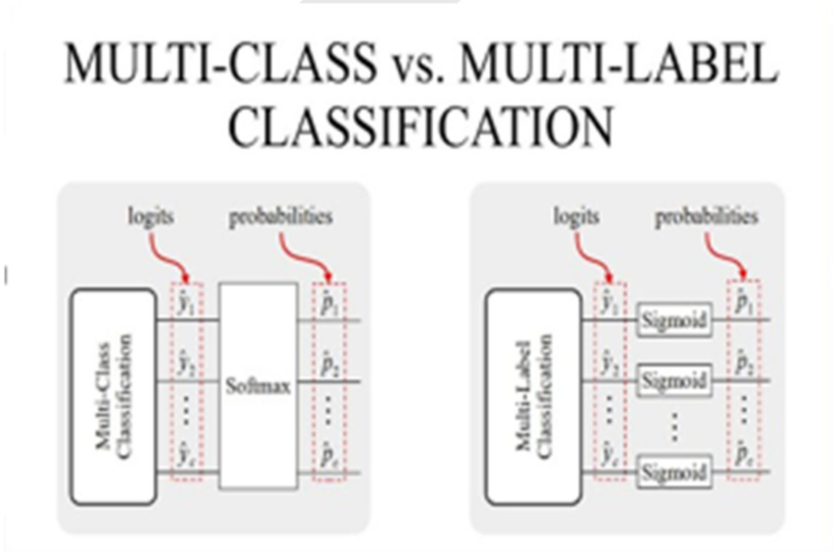
Analytical evaluation and benchmark testing

- Evaluate analytically means to systematically examine and interpret something using logic, reasoning, and careful analysis. This approach involves *breaking down complex ideas or problems into their component parts, understanding each part individually, and then synthesizing them to gain a comprehensive understanding of the whole*. In various contexts, such as mathematics, science, literature, or philosophy, evaluating analytically often involves:
  - **Breaking down**: Breaking a problem or concept into smaller, more manageable components or pieces.
  - **Examining closely**: Investigating each component thoroughly, understanding its properties, characteristics, and relationships with other components.
  - **Applying logic and reasoning**: Using logical reasoning, deduction, and inference to draw conclusions or make predictions based on the information gathered.
  - **Drawing conclusions**: Synthesizing the analyzed information to draw meaningful conclusions or make informed decisions.

- Benchmark testing refers to the process of comparing the performance of a system, device, program, or component against a standard set of metrics or against other similar systems or components. The primary goal of benchmark testing is to *assess and measure the performance, capabilities or efficiency of the item being tested under controlled conditions.*

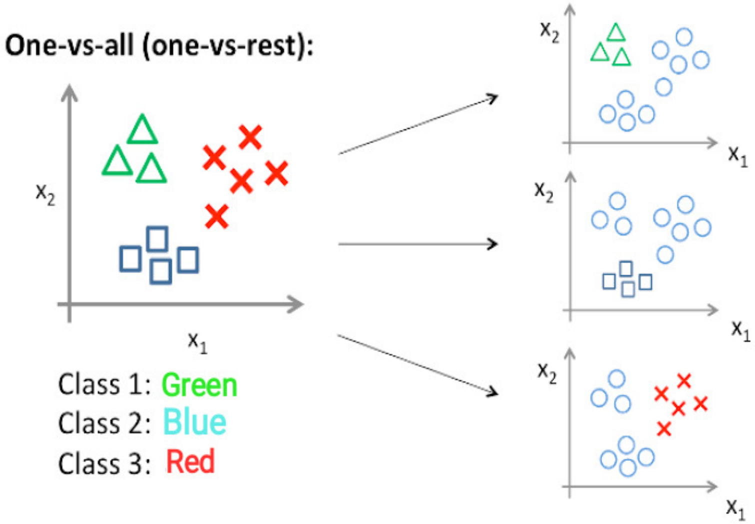
Klasifikavimas/klasterizavimas pagal daugelį žymių - Multi-lable classification/clusterring





Sources:
Madjarov, G., Kocev, D., Gjorgjevikj, D., and Džeroski, S. An extensive experimental comparison of methods for multi-label learning. *Pattern Recognition 45*, 9 (Sept. 2012), 3084–3104
https://youtu.be/Epx2V3Kd3dE

Classification strategies/klasifikavimo strategijos



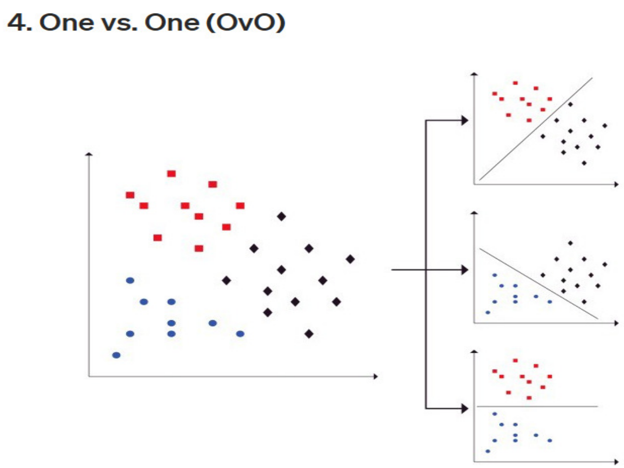
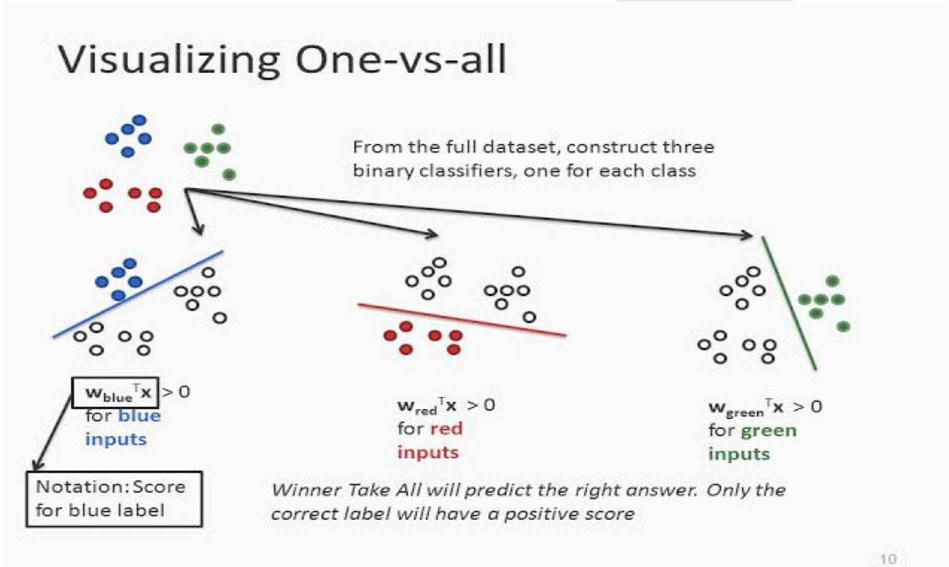


Figure 10: Photo via ScienceDirect.com

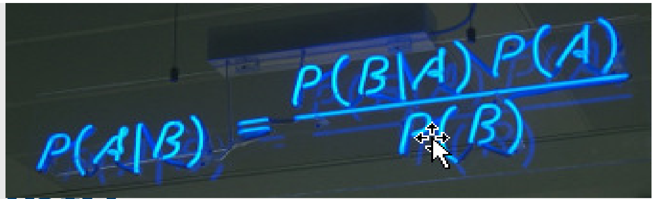


One vs all: generuojama N (=žymių skaičiui) dvejetainių klasifikatorių; analizuomos modelių prognozuotos tikimybės, rezulatu laikoma ta klasė, kuriai tikimybė didžiausia;

One vs one: generuojama N*(N-1) dvejetainių klasifikatorių-modelių; pirminė duomenų aibė dalinama į aibę kiekvienai klasei vs kiekvienos kitos klasės aibė;

Error correcting output code: generuojama N (=žymių skaičiui) dvejetainių klasifikatorių; kiekviena klasė reprezentuojama dvejetainiu kodu; kiekvienai klasei priskiriama xx bitų ilgio unikali dvejetainė eilutė (kodinis žodis). Apmokoma po vieną dvejetainį klasifikatorių kiekvienam stulpeliui; klasifikuojant naują egzempliorių x (data point), vertinami visi xx klasifikatorių - gaunama xx bitų eilutė; rezultate pasirenkama klasė, kurios kodinis žodis yra panašiausias į x.

Sources:
https://towardsdatascience.com/multi-class-classification-one-vs-all-one-vs-one-94daed32a87b ;
https://www.ccs.neu.edu/home/vip/teach/MLcourse/4_boosting/lecture_notes/ecoc/ecoc.pdf

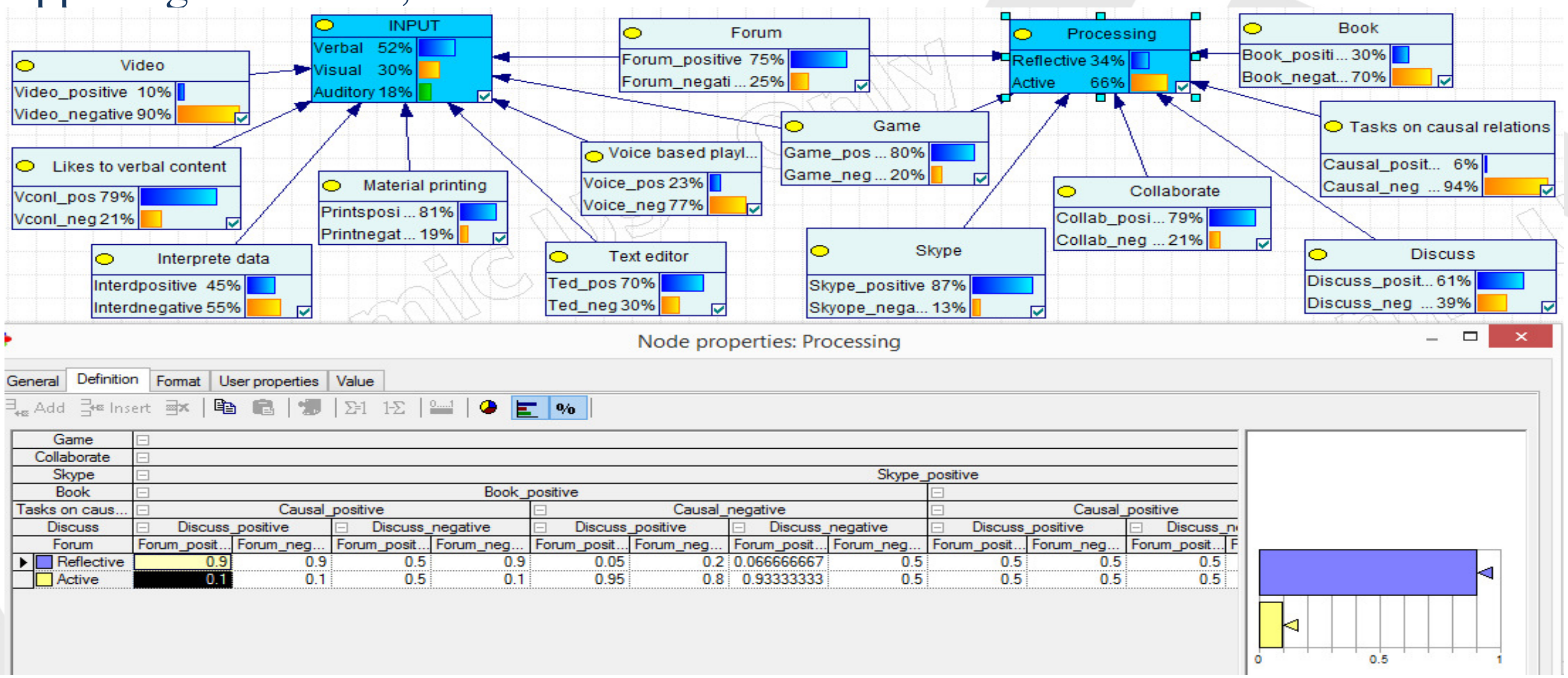


Bayesian networks: modelling uncertainty

- Bayesian networks aim to model conditional dependence, and therefore causation, by representing conditional dependence by edges in a directed graph. Through these relationships, one can efficiently conduct inference on the random variables in the graph through the use of influencing factors;
- the joint distribution for a Bayesian network is equal to the product of P(node|parents(node)) for all nodes, stated below:

$$P(X_1,\dots,X_n)=\prod_{i=1}^{n}P(X_i \mid X_1,\dots,X_{i-1})=\prod_{i=1}^{n}P(X_i \mid Parents(X_i))$$

- Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life;



Case-based reasoning (CBR)

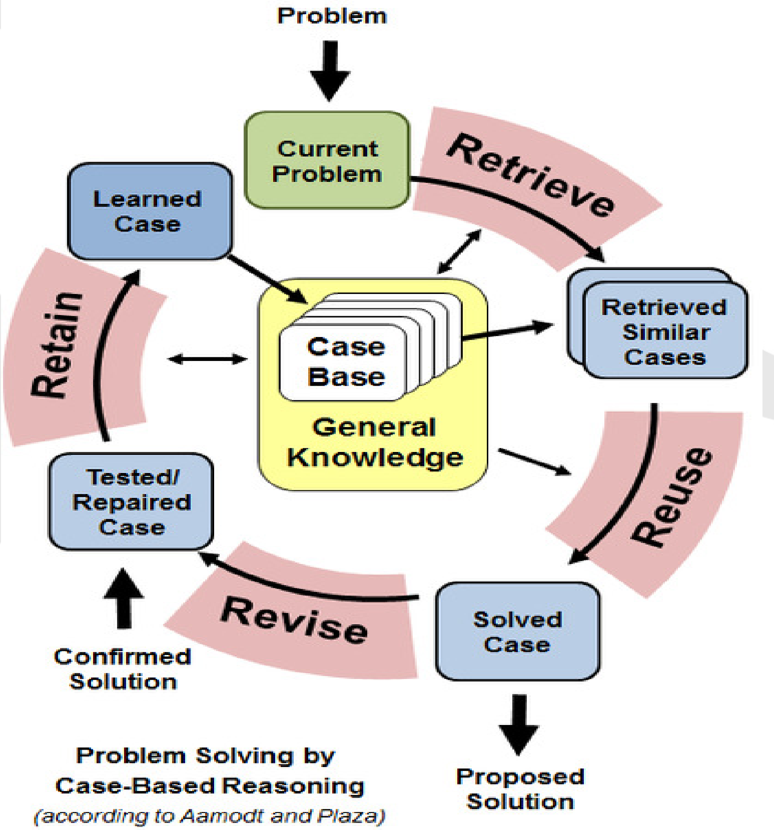
- a method which helps to solve new problems using knowledge from the past cases;
- divides an experience into two parts: a problem part (description of a problem situation) and a solution part (description of the reaction to a situation);
- to reuse cases from past, a recorded experience needs to be similar to the new problem. Typically, the new case is most similar to the nearest neighbour's case, therefore the global similarity must be computed:

$$\sum_{i=1}^{n}(w_i * sim(x_i, y_i) \mid 1 \leq i \leq n) .$$

BN: aleatory uncertainty (inherent randomness)

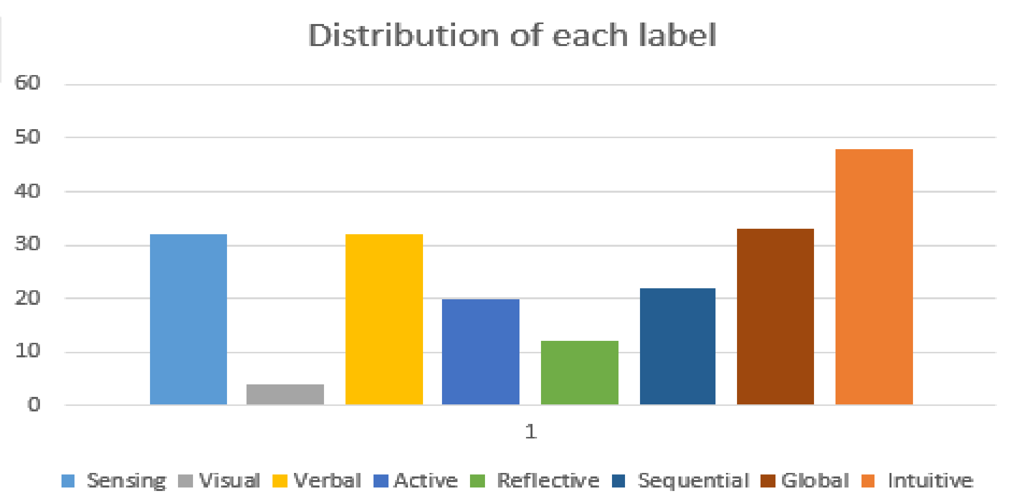
+

CBR: epistemic uncertainty (lack of knowledge, information or understanding)



Duomenų aibė

- Moodle duomenys (https://www.examulator.com/er/) nereprezentatyvūs, neinformatyvūs, nepakankami - sugeneruoti duomenys eksperimentavimui;
- sužymėta rankiniu būdu;
- 2 poros įvesties atributų stipriai koreliuoja;
- eksperimentai atlikti su:
  - ➢ subalansuota ir nesubalansuota duomenų aibėmis;
  - ➢ Kai duomenų aibėje žymės koreliuoja ir nekoreliuoja.



	Navigation_deep	Navigation_skip_overview	Forum_visit	Forum_post	Video_pictures	Content_text_stay	Feedback_no	NO_connections_links	Quiz_revisions	Ques_details	Ques_facts	Ques_concepts
Navigation_deep	1.000000	0.222226	-0.095902	-0.160932	0.072537	-0.041814	0.073805	0.060616	0.143638	0.108631	0.130398	0.028180
Navigation_skip_overview	0.222226	1.000000	-0.049857	-0.141393	0.135592	0.021228	0.037722	0.081192	0.046758	-0.012163	-0.014672	0.589941
Forum_visit	-0.095902	-0.049857	1.000000	0.741034	0.343269	0.274579	0.129579	0.123309	0.020828	0.061617	0.002406	-0.184271
Forum_post	-0.160932	-0.141393	0.741034	1.000000	0.176030	0.667741	0.107686	0.126812	0.030363	0.348158	0.015333	-0.155371
Video_pictures	0.072537	0.135592	0.343269	0.176030	1.000000	0.288444	0.667107	0.601353	0.288067	0.190644	0.258271	0.081917
Content_text_stay	-0.041814	0.021228	0.274579	0.667741	0.288444	1.000000	0.288288	0.283076	0.169344	0.603243	0.167712	0.065209
Feedback_no	0.073805	0.037722	0.129579	0.107686	0.667107	0.288288	1.000000	0.865460	0.184174	0.185305	0.187137	-0.046124
NO_connections_links	0.060616	0.081192	0.123309	0.126812	0.601353	0.283076	0.865460	1.000000	0.221242	0.221631	0.150227	-0.031586
Quiz_revisions	0.143638	0.046758	0.020828	0.030363	0.288067	0.169344	0.184174	0.221242	1.000000	0.684562	0.925804	-0.026526
Ques_details	0.108631	-0.012163	0.061617	0.348158	0.190644	0.603243	0.185305	0.221631	0.684562	1.000000	0.598606	-0.013862
Ques_facts	0.130398	-0.014672	0.002406	0.015333	0.258271	0.167712	0.187137	0.150227	0.925804	0.598606	1.000000	-0.041687
Ques_concepts	0.028180	0.589941	-0.184271	-0.155371	0.081917	0.065209	-0.046124	-0.031586	-0.026526	-0.013862	-0.041687	1.000000

Assignment	Analytics	Badges	Book
Chat	Choice	Course	Data
Enrolment	Feedback	Forum	Lesson
LTI	Page	Quiz	Scorm
Survey	Users	Workshop	Wiki

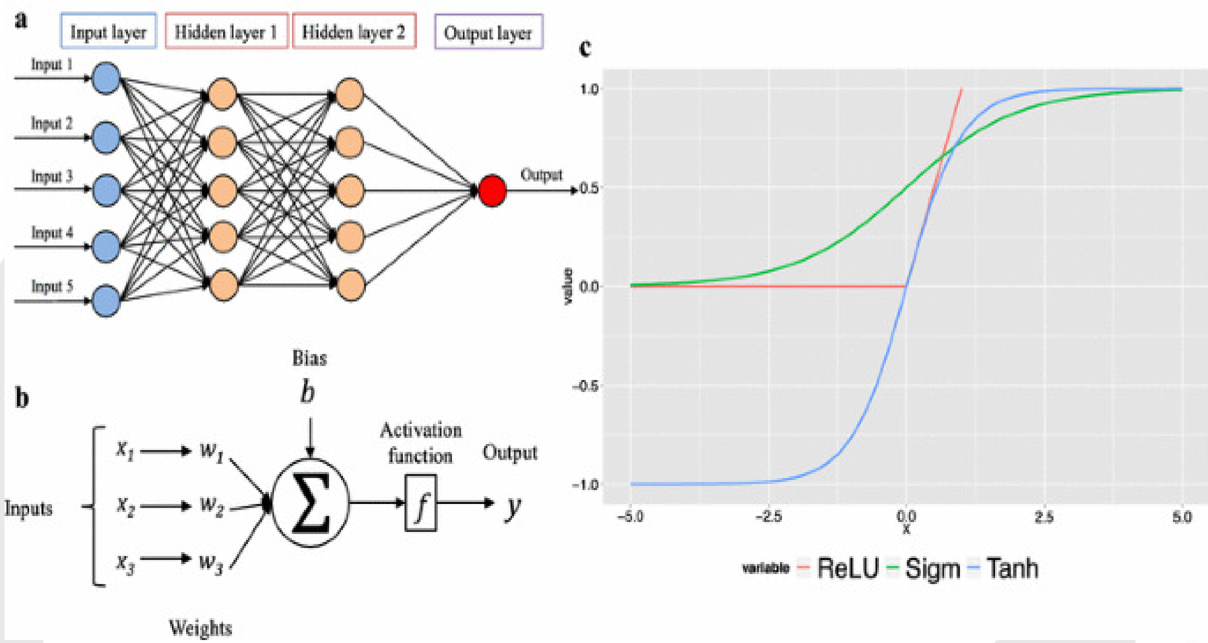
Application of problem transformation and ensemble methods/problemos transformavimo ir ansamblio metodų taikymas

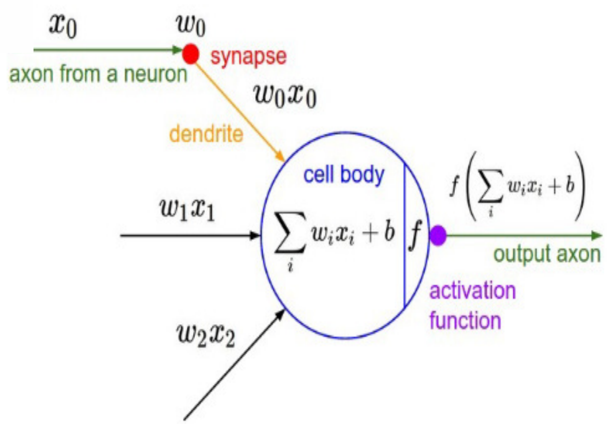
- Baziniai estimatoriai: *MultinomialNB, SGDClassifier, LogisticRegression, LinearSVC, Perceptron, GradientBoostingClassifier, PassiveAggressiveClassifier*
- ***Binary relevance*** (For each label, a separate binary classifier (e.g., logistic regression, SVM, decision tree) is trained to predict whether an instance belongs to that label or not. Each classifier is trained independently of the others; the classifiers for each label are trained independently, ignoring any correlations or dependencies between labels; During prediction, each binary classifier outputs a probability or confidence score for its respective label. The final prediction for each label is determined independently based on these scores), ***Label Powerset*** (each unique combination of labels in the dataset is treated as a separate class; transforms the multi-label classification problem into a multi-class classification problem, where the goal is to predict the correct combination of labels); ***ensemble*** – koreliuojančioms žymėms;
- ***OneVsOne*** (deals with mutually exclusive choices; a binary classifier is trained for each pair of labels in the dataset; each binary classifier distinguishes between one pair of labels; during prediction, each binary classifier votes for one label, and the final prediction is determined by the label that receives the most votes across all binary classifiers) ir ***OneVsRest*** strategijų pritaikymas;

One Vs Rest

- to use OvR for multi-label classification, it is needed to transform the problem into *m* independent binary classification tasks, where *m* is the number of unique labels in the dataset; for each unique label, a separate binary classifier (e.g., logistic regression, SVM, decision tree) is trained to predict whether an instance belongs to that label or not. Each binary classifier is trained independently of the others. During prediction, each binary classifier is applied to the instance to obtain a probability or confidence score for each label. The decision threshold may be used to determine whether a label should be assigned to the instance or not. For example, if the confidence score exceeds a certain threshold (e.g., 0.5), you can consider the label as predicted for that instance.
- Advantages:
  - better for correlated labels;
  - OvR creates balanced binary classification problems for each class, potentially mitigating the impact of class imbalance;
  - computationally efficient;
  - simpler to implement;
  - scalable, especially in scenarios with a large number of labels. The number of binary classifiers in OvR scales linearly with the number of classes(BR – exponentially);
- Disadvantage:
  - assumes that each class is mutually exclusive, which may not always be the case in real-world scenarios.

Experimenting with Neural network – eksperimentai su neuroniniais tinklais [18]





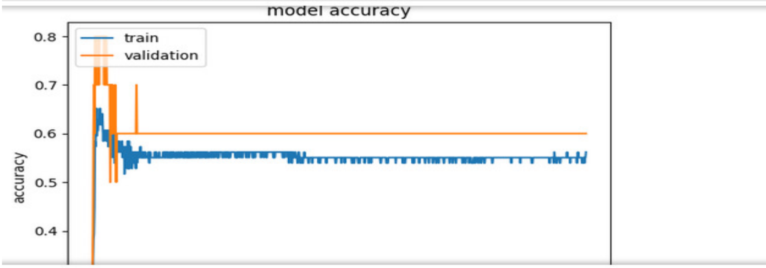
Experimenting with Neural network – eksperimentai su neuroniniais tinklais

- Grid search: batch_size – 10, number of epochs – 1000; learning rate is 0.01; 25 neuronai; momentum-0.4;

```
[67]: # get the model
def get_model(n_inputs, n_outputs, neurons):
    model = Sequential()
    model.add(Dense(neurons, input_dim=n_inputs, kernel_initializer='he_uniform', activation='relu'))
    model.add(Dense(12, kernel_initializer='he_uniform', activation='relu'))
    model.add(Dense(n_outputs, activation='sigmoid'))
    model.compile(loss='binary_crossentropy', optimizer='adam',metrics=["accuracy"])
    model.save('C:/Users/Daiva1/Desktop/model.h5')
    model_final = load_model('C:/Users/Daiva1/Desktop/model.h5')
    return model_final
```

- The average model accuracy obtained was 0.45 - the model predicted the exact label 45.4 % of the time):

```
# evaluate model
results = evaluate_model(X, y)
```




```
# summarize performance
#accuracy SCORE = number of correct predictions (If normalize=False)- for all labels!
print('Model evaluation by accuracy score: %.3f (%.3f)' % (mean(results), std(results)))
Model evaluation by accuracy score: 0.454 (0.163)
```

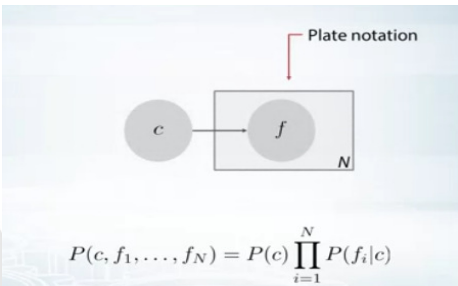
- imbalanced data set, but without applying imbalanced weight method: accuracy score was *0.496* ; imbalanced data set, applying imbalanced weight method: 0.43;

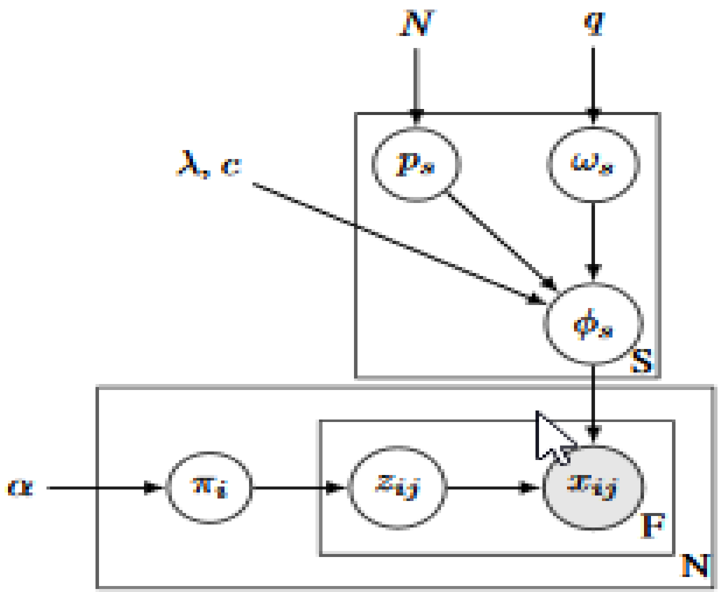
With imbalanced weights methods, the loss function is modified to assign different weights to different classes. These weights are typically inversely proportional to the class frequencies. The minority class is given a higher weight, while the majority class is given a lower weight. By assigning higher weights to the minority class, the algorithm places more emphasis on reducing errors in predicting this class correctly. This means that misclassifications of the minority class contribute more to the overall loss, and the algorithm is incentivized to improve its performance on this class. By adjusting the class weights, the algorithm effectively balances the importance of different classes in the training process. This helps mitigate the bias towards the majority class that can occur in imbalanced datasets. As a result, the trained model is more likely to make accurate predictions for both the majority and minority classes. Adding class weight but not changing the way the model performance is measured can degrade the overall performance of the model since it is designed to allow increased loss on lower-weighted classes. The model may perform even worse overall but the performance is better on the heavily weighted classes.

BN+CBR modelling [1] [2]

- examplar-based: doesn't use pre-defined learning style model – model is generated from his/her behavioral data or students' past cases' data; result – distribution of features over clusters;
- example – Bayesian Case model (BCM makes joint inference on cluster labels, prototypes and subspaces) - (prognozuoja mokymosi stilių klasterių proporcijas);

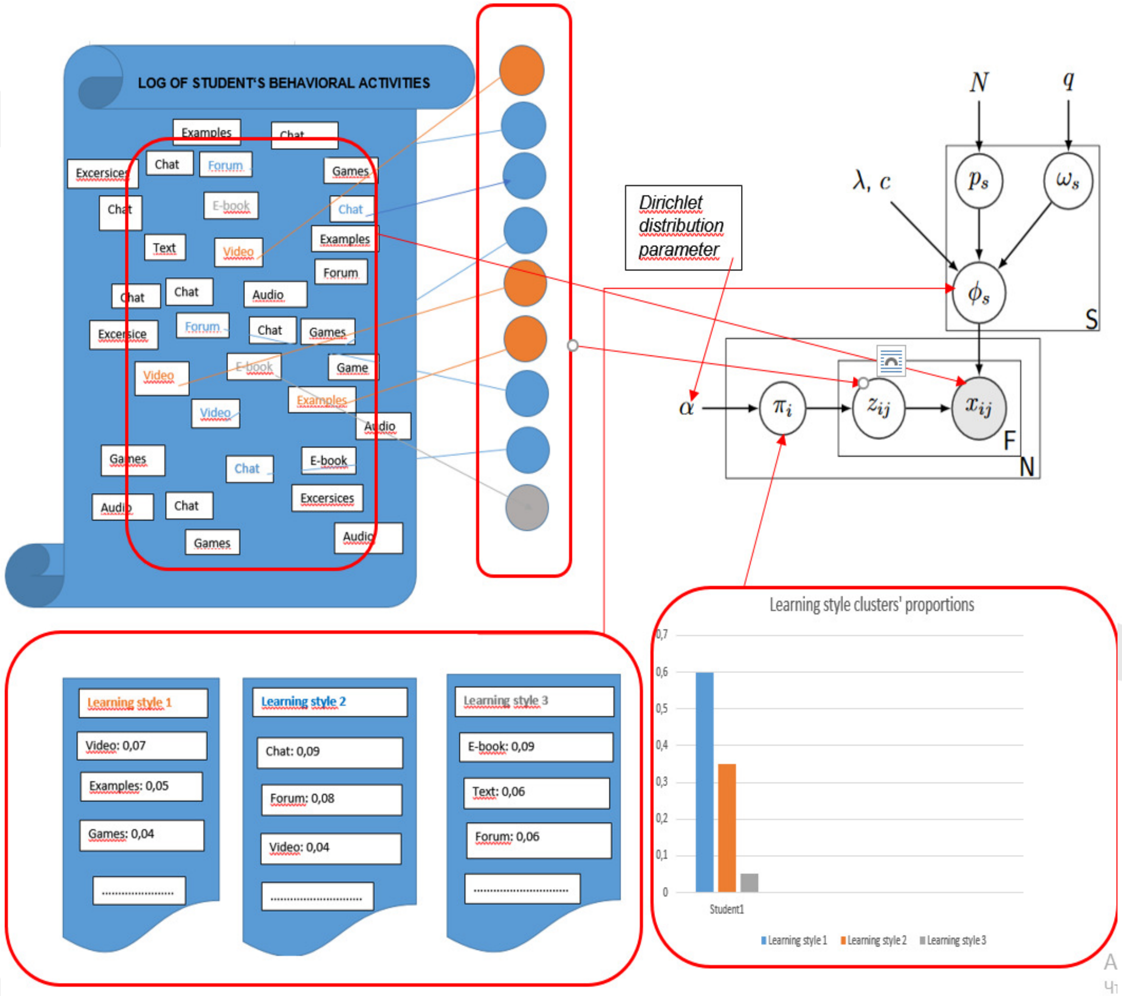
- $\phi_{sj} \sim \text{Dirichlet}\left(g_{psj,wsj,\lambda}\right) \forall\, s, j$; where $g_{psj,wsj,\lambda}(v) = \lambda(1 + c\mathbf{1}_{[w_{sj}=1 \text{ and } p_{sj}=\vartheta]})$;
- $p_s \sim \text{Uniform}(1, N) \forall\, s$;
- $w_{sj} \sim \text{Bernoulli}(q) \forall\, s, j$;
- $\pi_i \sim \text{Dirichlet}(\alpha) \forall\, i$;
- $z_{ij} \sim \text{Multinomial}(\pi_i) \forall\, i, j$;
- $x_{ij} \sim \text{Multinomial}\left(\phi_{z_{ij}j}\right) \forall\, i, j$;





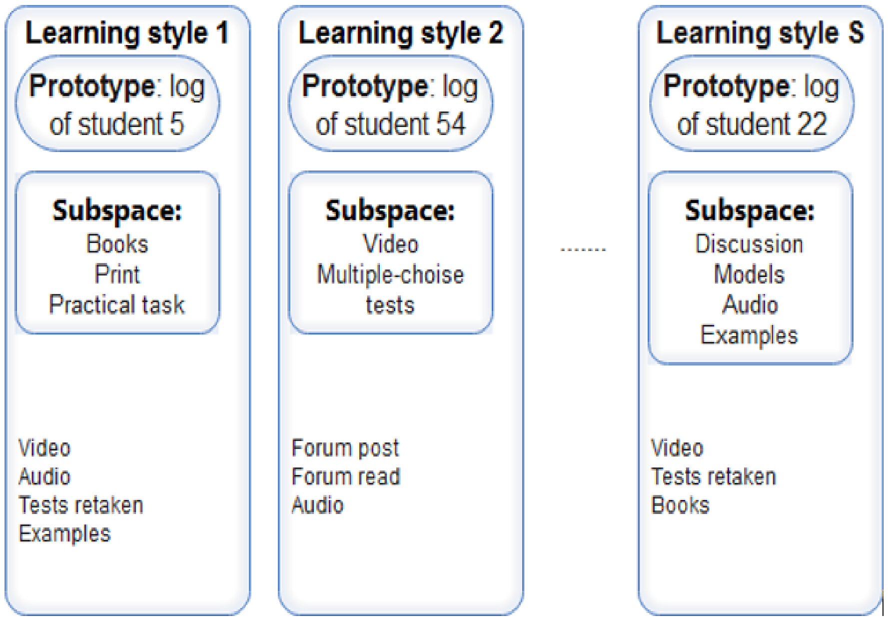
Bayesian case model (BCM) for student learning style modelling [1] [2]

- **generative model** which aims to capture the underlying probability distribution of the observed data. By learning the parameters of this distribution, the generative model can generate new samples that are statistically similar to the original data;
- incorporate latent variables, which are unobserved variables that capture underlying factors of variation in the data. By modeling the joint distribution of observed and latent variables, these models can learn a more compact representation of the data and generate new samples by sampling from the latent space;
- trained using maximum likelihood estimation or Bayesian inference. Maximum likelihood estimation involves maximizing the likelihood of the observed data given the model parameters, while Bayesian inference involves estimating the posterior distribution of the model parameters given the observed data and prior knowledge;
- once trained, generative models can be used for both sampling and inference tasks. Sampling involves generating new samples from the learned probability distribution, while inference involves estimating the latent variables or model parameters given observed data.



BCM clusters, subspaces, prototypes for learning style modeling [1] [2]

- interpretable models and methods for interpretation enable humans to comprehend why certain decisions or predictions have been done;
- interpretability means the degree to which a human can consistently predict the model's result;
- one can describe a model as interpretable if he/she can comprehend the entire model at once;



- explanation for each cluster will consist of:
  - ✓ prototype presented as log of behavioral activities of student which best represents the cluster (or best represents learning style assigned to the cluster);
  - ✓ subspace of important features, i. e. behavioral activities that have been performed most frequently in the virtual hypermedia learning environment by students whose learning style corresponds to the style represented by the cluster;

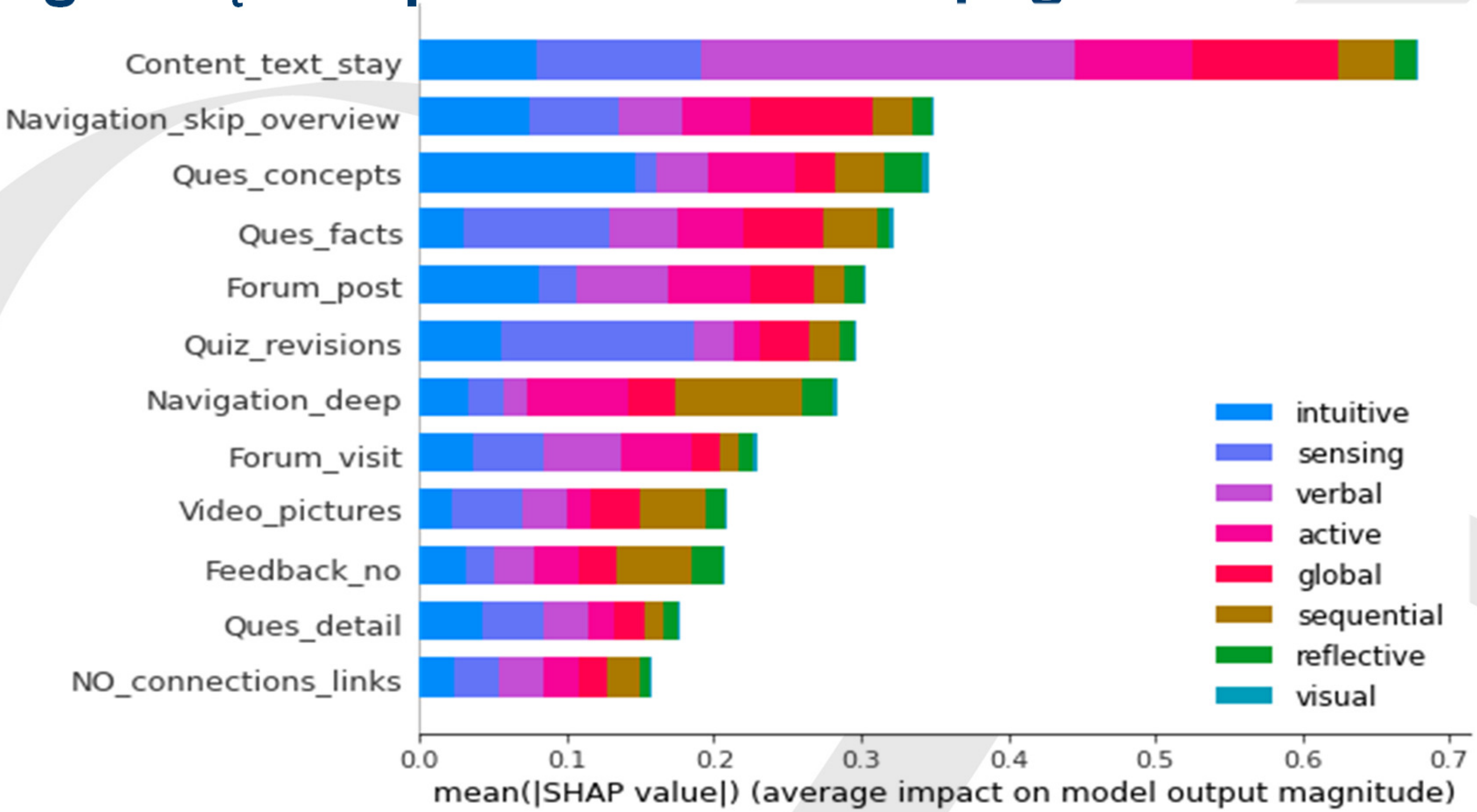
MCMC ir Gibso atrankos (GA) algoritmas (nuoseklus)

- **Įvestis**: sąlyginis pasiūlymas (angl. conditional proposal)
- **Išvestis:** sugeneruotų iš P(x) n koreliuotų egzempliorių aibė
- **Inicializuoti** būseną $<x_1^{(0)}, \dots, x_n^{(0)}>$
- **Inicializuoti** generuojamą aibę kaip Null (tuščią)
- **For** k=1…N kartoti
-     **For** i=1…n kartoti
-         Generuoti $x_i^{(k+1)} \sim P\left(x_i^{(k)} \middle| x_1^{(k)}, \dots, x_{i-1}^{(k)}, x_{i+1}^{(k)}, x_n^{(k)}\right)$
-         Pridėti $x_i^{(k+1)}$ į sugeneruotų egzempliorių aibę
- **Pabaiga For**
- **Pabaiga For**
- MCMC: perėjimo į sekančią būseną tikimybė priklauso tik nuo einamosios būsenos ir nepriklauso nuo ankstesnių būsenų; Markov chains can be used to model sequences of data. Markov chains can be used to generate sequences of data by sampling from the transition probabilities between states. Each state in the Markov chain corresponds to an element in the sequence, and the transition probabilities determine the likelihood of transitioning from one element to the next. Transition probabilities represent the likelihood of transitioning from one word to another based on the observed frequencies of word co-occurrences in the training text. By iteratively sampling from the transition probabilities, a Markov chain can generate sequences that resemble the patterns observed in the training data.
- GA: nevykdome atrankos tiesiogiai iš aposteriorinio skirstinio, bet modeliuojame kiekvieno latentinio kintamojo/parametro tikimybių reikšmes iš aposteriorinio sąlyginio skirstinio; Gibbs sampling is used for generating samples from high-dimensional probability distributions. It is particularly useful when the joint distribution of the variables is known, but sampling from the joint distribution directly is difficult or inefficient.
- GA: algoritmas yra inicializuojamas, atsitiktiniu būdu parenkant pradines reikšmes;

Defended statements/Ginamieji teiginiai

1. Since more than one learning style dimension usually characterizes a student, to predict the probabilities of all learning style dimensions, problem transformation multi-label classification method applying *OneVsRest* approach is preferred over the Binary relevance method in case of imbalanced data sets and/or correlated labels. *OneVsRes*t approach enables to interprete the prediction results since each class is represented by one and only one classifier and it is possible to gain knowledge about the class by inspecting its corresponding classifier. *MultinomialNB* or *GradientBoostingClassifier* should be preferred as base estimators.
2. In supervised learning, for multi-label classification of student's learning style, *Label Powerset* classification outperformes the *Binary relevance* classification. *Label Powerset* considers correlation between labels indirectly, and *Binary relevance* ignores correlations between labels.
3. When labels are correlated, a chain of binary classifiers should be constructed, wherein a classifier Ci uses the predictions of all the classifiers, Cj, where $j < I$, or the *Label Powerset* method have to be chosen.
4. For the imbalanced data sets, *imbalanced weights* methods that are designed to allow increased loss(error rate, the price paid for inaccuracy) on lower-weighted classes is applicable. In the latter case, the model may perform even worse overall than before applying imbalanced weights method, but the performance is better on the heavily weighted classes. In case of the application of problem transformation methods with imbalanced data sets, *One vs One* approach is recommended over the *Label Powerset* method since it avoids generalization. Tree-based methods also may be a good option.
5. Unsupervised learning methods that use generative approach are recommended in case there are no labeled data and the groups of students having different learning styles are not well-separated and overlapping. In order to account for influences of aleatory and epistemic uncertainties in student learning style modelling, it is reasonable to combine CBR method and Bayesian Case model. Exemplar-based generative learner modelling (e. g. BCM) is a good choise in cases when an inherent interpretability by example is needed.
6. Interpretability (supervised learning, multi-label classification task): when behavioural activities of the learner (the input features) are not strongly correlated, the SHAP model-agnostic method is the best option.

Shapley values for interpretability of the model/modelio prognozių interpretavimas SHAP pagalba

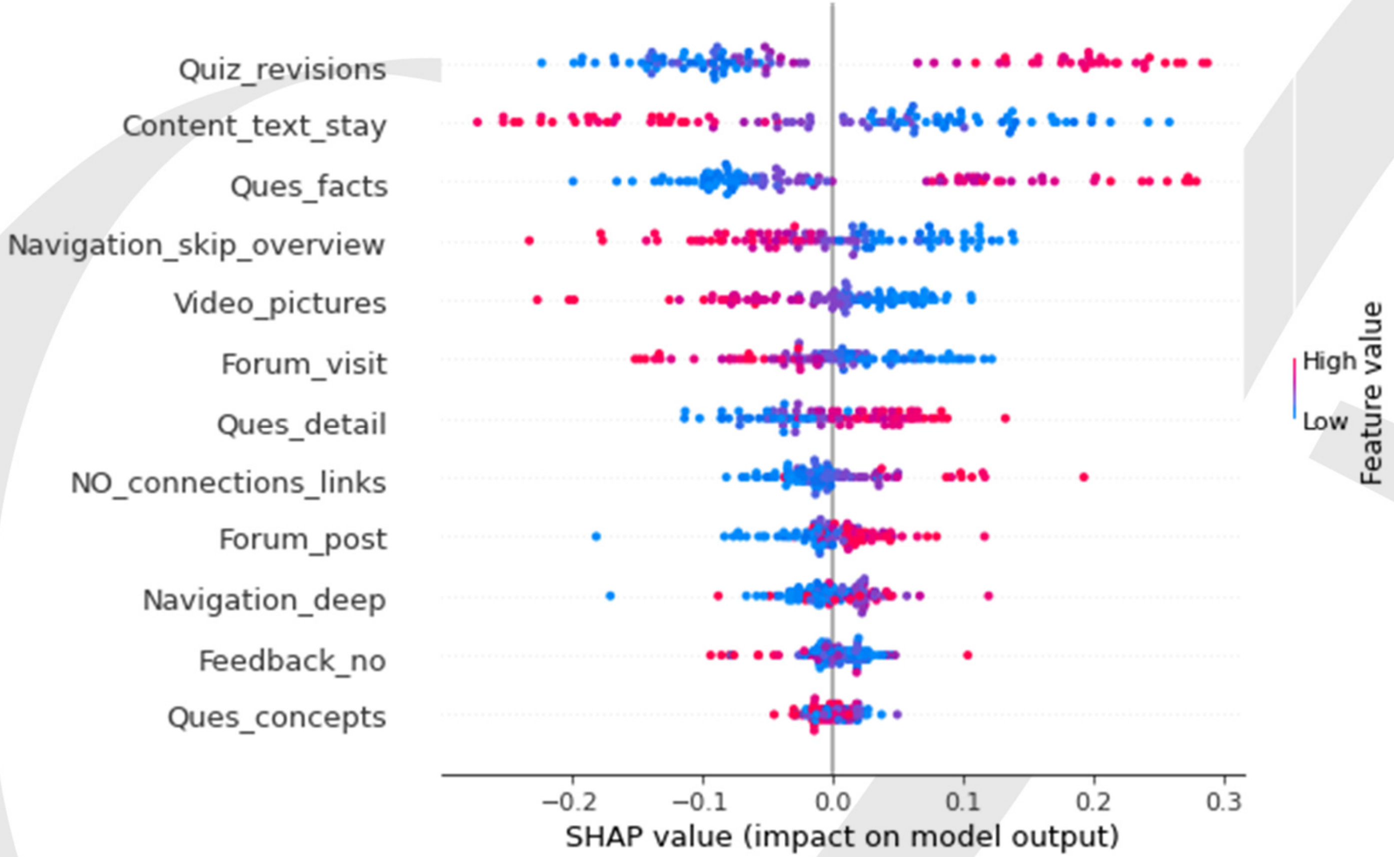


Shap values show how much a given feature changed our prediction (compared to if we made that prediction at some baseline value (pradinė reikšmė) of that feature) - quantify the marginal contribution of each feature to the difference between the actual prediction and the average prediction made by the model across all possible subsets of features.

The value of the j-th feature contributed ϕj to the prediction of this particular instance compared to the average prediction for the dataset. (x- feature value contribution)

Features with large absolute Shapley values are important. If we want the global importance, we average the ***absolute*** *Shapley values per feature across the data.*

Example of the shap values for the first label (https://christophm.github.io/interpretable-ml-book/shap.html)



High values of a feature (indicated by rose/purple combination) - leads to prediction 1; low values of feature (indicated by blue) - leads to prediction 0;

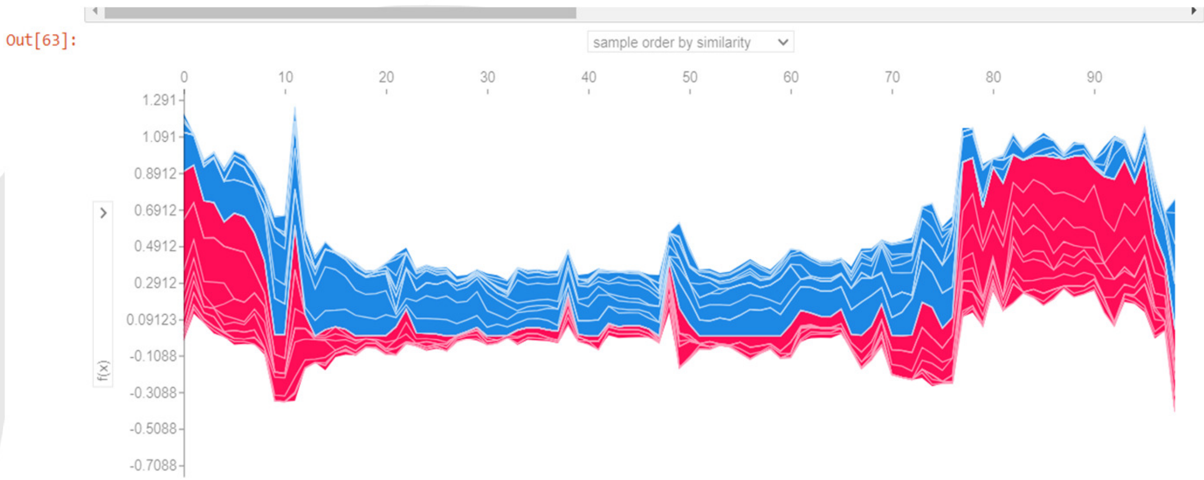
SHAP summary plot combines feature importance with feature effects: each point on the summary plot is a Shapley value for a feature and an instance. The position on the y-axis is determined by the feature and on the x-axis by the Shapley value. The color represents the value of the feature from low to high.

Contribution of features to the lable prediction – kintamųjų reikšmių indėlis į prognozę

The output value is the prediction for that observation (the prediction of the first row in Table B is 6.20).
 # The base value: The original paper explains that the base value E(y_hat) is "the value that would be predicted if we
 # did not know any features for the current output." In other words, it is the mean prediction, or mean(yhat).
 # You may wonder why it is 5.62. This is because the mean prediction of Y_test is 5.62.
 # You can test it out by Y_test.mean() which produces 5.62.
 # Red/blue: Features that push the prediction higher (to the right) are shown in red, and those pushing the prediction lower are in blue.

The baseline – the average predicted probability

shap.force_plot(explainer.expected_value[1], shap_values[1], feature_names = ...



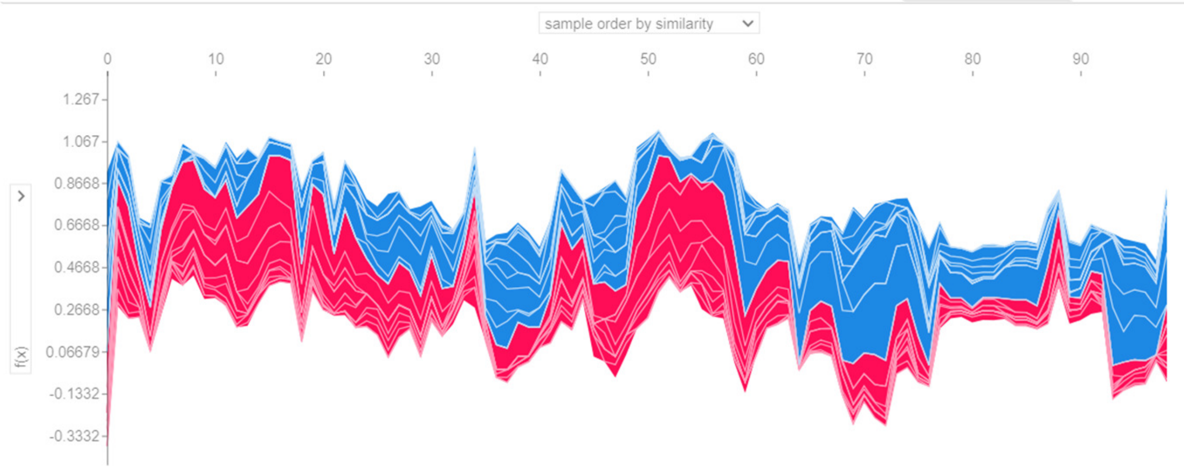
Stacked SHAP explanations clustered by explanation similarity. Each position on the x-axis is an instance of the data. Red SHAP values increase the prediction, blue values decrease it.

Galima pritaikyti studentų/logų klasterizavimui.

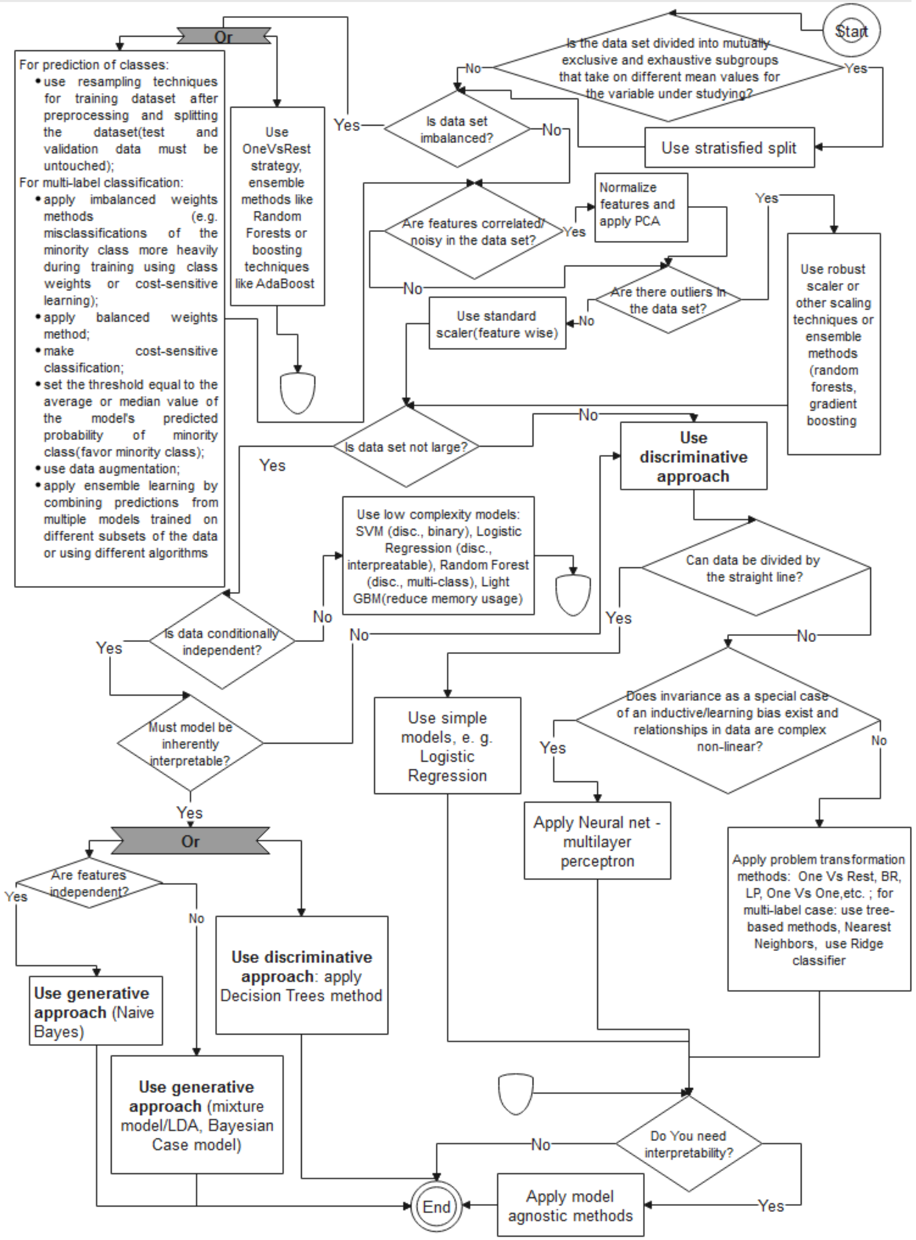
The force plot shows how features explain the model output for multiple observations at the same time.

We can cluster the data with the help of Shapley values. The goal of clustering is to find groups of similar instances. Normally, clustering is based on features. SHAP clustering works by clustering the Shapley values of each instance. This means that you cluster instances by explanation similarity.

This plot (interactive in the notebook) is the same as individual force plot. Just imagine multiple force plots rotated 90 degrees and added together for each example.

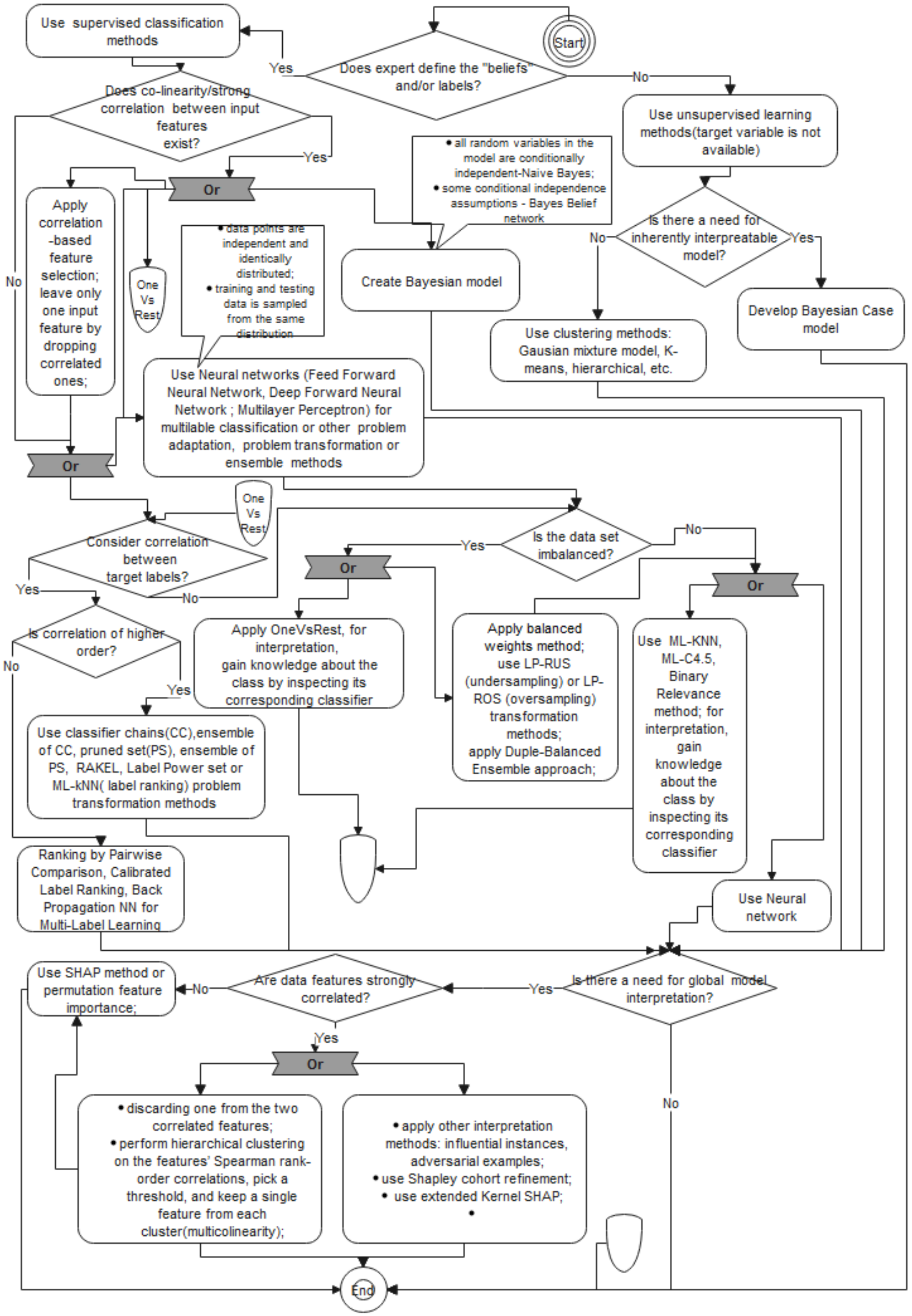


Procedure for selecting ML methods for learning style detection model



VILNIUS GEDIMINAS
TECHNICAL UNIVERSITY
FACULTY OF FUNDAMENTAL SCIENCES

Procedure for selection of methods for multi-label students’ learning style detection



```
[23]: # summarize performance
      #accuracy = number of correct predictions / total predictions
      print('MAE: %.3f (%.3f)' % (mean(results), std(results)))

      MAE: 0.266 (0.119)
```

Validation/Validavimas

- Multi-label classification data set of 100 000 samples was generated (labels, features are not strongly correlated);
- Grid search applied for neural network:
- NN was trained with optimal hyper parameters, using imbalanced weights method – accuracy score obtained is ;
- prediction on unseen data:

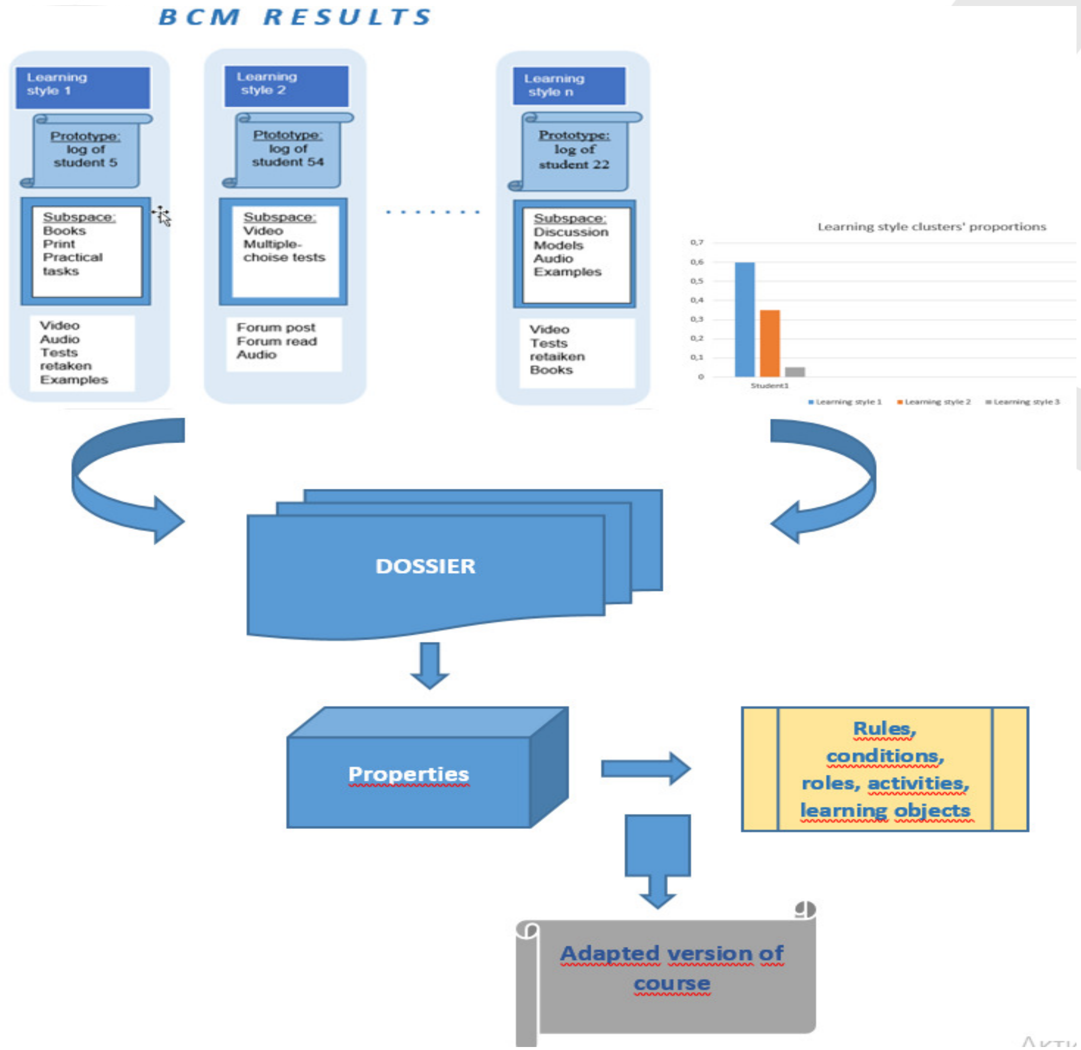
```
ones_count_per_column=np.sum(y, axis=0)
print(ones_count_per_column)

1Sensing        84382
2Intuitive      89263
3Visual         54806
4Verbal         86496
5Active         88832
6Reflective     43913
7Sequential     86808
8Global         76984
dtype: int64
```

```
# fix random seed for reproducibility
from keras.models import load_model
from scikeras.wrappers import KerasClassifier
seed = 7
tf.random.set_seed(seed)
# create model
model=KerasClassifier(model=get_model, n_inputs=12, n_outputs=8, verbose=0)
# define the grid search parameters
#batch_size = [20, 40, 60, 80, 100]
epochs = [1000]
learn_rate = [0.001, 0.01]
neurons = [15, 25]
#batch_size=batch_size, epochs=epochs,optimizer__learning_rate=learn_rate, optimizer__momentum=momentum,
#param_grid = dict( model__neurons=neurons)
#momentum = [0.0, 0.2, 0.5, 0.8]
param_grid = dict(epochs=epochs,  model__neurons=neurons,optimizer__learning_rate=learn_rate)
#, optimizer__learning_rate=learn_rate, optimizer__momentum=momentum, model__neurons=neurons), optimizer__momentum=momentum
grid = GridSearchCV(estimator=model, param_grid=param_grid, refit="accuracy_score", pre_dispatch=1, error_score="raise")
grid.set_params(n_jobs=1, cv=3,scoring=scorers)
grid_result = grid.fit(X, y)
grid_result.cv_results_
# summarize results
print("Best: %f using %s" % (grid_result.best_score_, grid_result.best_params_))
# means = grid_result.cv_results_['mean_test_score']
# stds = grid_result.cv_results_['std_test_score']
# params = grid_result.cv_results_['params']
# for mean, stdev, param in zip(means, stds, params):
    # print("%f (%f) with: %r" % (mean, stdev, param))

Best: 0.297510 using {'epochs': 1000, 'model__neurons': 25, 'optimizer__learning_rate': 0.001}
```

Integration of learning style into virtual learning environment – studento mokymosi stiliaus integravimas į virtualią mokymosi aplinką



Scientific novelty of the dissertation

- to predict all student's learning style dimensions, especially in scenarios with correlated labels, non-linear decision boundaries and/or class imbalances, multi-label classification approach applying *OneVsRest* strategy was proposed and the models applying it have been developed for various base estimators – multi-class approaches that treat each label separately were basically applied so far;
- models for multi-label learning style prediction were developed, both modelling and prediction dynamics were captured and prototypes of the models validated;
- since existing machine learning frameworks are too general and the existing machine learning cheat sheets focus rather on individual algorithms than on the entire solution to a problem and possible cases of the solution, the unmet needs were addressed by creating an original methodological freamework for selection and/or adjustment of methods for student‘s multidimentional learning style modelling under different conditions and/or restrictions was developed. It integrates insights and methods from multiple disciplines and provides a holistic, comprehensive and accurate understanding of a complex learning style modelling problem. Framework includes consistent procedures that may be used for theorethical and practical learner modelling;
- novel use case of the combination of Bayesian network and case-based reasoning method was proposed (aleatory(randomness in data)+epistemic(lack of training data);
- both the train of thought and know-how for the integration of a learner model into VLE were proposed. The approach presents a way of integration of the predicted individual learning style of a student or learning style of a group of students into VLE for personalization purpose.

Aspects of novelty and originality of cheat sheets

Cheat sheets, often used as quick reference guides or summaries, can indeed possess aspects of novelty or originality despite their concise nature. Here are several ways in which a cheat sheet can exhibit novelty or originality:

- **Unique Content Organization**: A cheat sheet may present information in a novel way, such as arranging concepts hierarchically, categorically, or through visual diagrams that haven't been commonly used before. Unconventional organization can help users grasp the material more effectively.
- **Innovative Visual Design**: Originality can stem from the visual design of the cheat sheet. Creative use of typography, colors, icons, or other graphical elements can make the cheat sheet more engaging and memorable.
- **Distinctive Content Selection**: Including lesser-known or advanced concepts alongside basic ones can add novelty to a cheat sheet. It can provide value to users who seek more comprehensive resources or those looking for a deeper understanding of the topic.
- **Novel Tips and Tricks**: Providing unique tips, shortcuts, or mnemonic devices can set a cheat sheet apart. These can help users remember key information more easily or offer alternative approaches to problem-solving.
- **Personal Insights or Annotations**: Adding personal insights, anecdotes, or annotations to the cheat sheet can inject originality. This might include explanations of why certain concepts are important, common mistakes to avoid, or real-world examples.
- **Customization Options**: Offering customization options, such as fillable fields or blank spaces for users to add their own notes, can make the cheat sheet more versatile and adaptable to individual needs.
- **Interactive Elements**: Incorporating interactive elements, such as clickable links, pop-up definitions, or interactive quizzes, can enhance user engagement and provide a unique learning experience.
- **Collaborative Creation**: Collaboratively creating a cheat sheet with contributions from multiple experts or users can result in a document with diverse perspectives and insights, adding to its originality.
- **Tailored to Specific Audiences**: Designing cheat sheets specifically for niche audiences or specialized topics can make them more original and relevant. For example, a cheat sheet tailored for data scientists may include unique algorithms or statistical methods.
- **Integration of Emerging Trends**: Including information on emerging trends, technologies, or methodologies can keep the cheat sheet up-to-date and relevant, adding a layer of novelty.

By incorporating one or more of these aspects, a cheat sheet can stand out as a valuable and original resource for learners and practitioners in various fields.

Originality aspects

- **Innovative Approach**: If the modeling framework introduces **a novel methodological approach** or technique that has not been previously utilized in the field, it can be considered original. This could involve the application of advanced mathematical or computational methods, the integration of diverse data sources, or the adaptation of existing models in a novel context.
- **Addressing Unmet Needs**: **If the modeling framework addresses a gap or limitation in existing approaches by providing a more comprehensive or accurate representation of the phenomenon under study, it can be considered original**. This could involve improving the predictive accuracy of models, enhancing their explanatory power, or enabling the exploration of new research questions that were previously inaccessible.
- **Interdisciplinary Integration**: Originality can also stem from the integration of insights, methodologies, or data from multiple disciplines to develop a more holistic understanding of a complex system or problem. By bridging disciplinary boundaries, the modeling framework may offer new perspectives or solutions that have not been explored before.
- **Empirical Validation**: If the modeling framework is empirically validated through comparison with real-world data or experimental results, demonstrating its utility and effectiveness in capturing the dynamics of the system under study, it can be considered original. Validation provides evidence of the framework's practical relevance and contributes to its credibility within the scientific community.
- **Application in Novel Contexts**: Originality can also arise from applying a modeling framework developed for one context to a novel domain or problem area, where it yields new insights or facilitates the exploration of previously uncharted territory. This demonstrates the versatility and adaptability of the framework beyond its original scope.
- Overall, the originality of a modeling framework lies in its ability to offer new perspectives, insights, or solutions that **advance the state of knowledge in a particular field, whether through methodological innovation, addressing unmet needs, interdisciplinary integration, empirical validation, or application in novel contexts.**

Novelty/originality aspects of developing a framework

The novelty aspects of developing a framework typically involve introducing new ideas, approaches, or perspectives that advance the field or address previously unexplored problems. Originality aspects characteristic to a newly developed modeling framework often revolve around introducing **customization options, community building** efforts, etc.

- Innovative Conceptualization: Introducing novel theoretical constructs or conceptual frameworks that offer fresh perspectives on understanding a phenomenon or solving a problem. This could involve synthesizing existing theories in a new way or proposing entirely new theoretical frameworks.
- Integration of Diverse Disciplines: Bringing together insights from different disciplines or fields to create a more comprehensive framework. Interdisciplinary frameworks can offer novel solutions to complex problems by leveraging diverse perspectives and methodologies.
- Domain-Specific Solutions: Tailoring the modeling framework to specific domains or applications by incorporating domain-specific knowledge, data preprocessing techniques, or specialized model architectures. Providing out-of-the-box support for particular industries or use cases can differentiate the framework from generic solutions.
- Addressing Understudied Areas**:** **Filling gaps in existing knowledge by developing frameworks for understudied areas** or marginalized communities. **This involves considering perspectives and experiences that have been historically overlooked in the field.**
- Methodological Innovations: Introducing novel research methods or analytical techniques within the framework development process. This could include using advanced statistical methods, computational modeling, or qualitative approaches to enhance the framework's robustness and applicability.
- Scalability and Generalizability: Designing frameworks that are scalable and adaptable to different contexts or levels of analysis. Novel approaches to scalability ensure that the framework remains relevant and effective across diverse settings and populations.
- User-Centered Design: Incorporating principles of user-centered design to develop frameworks that are intuitive, user-friendly, and tailored to the needs of practitioners or stakeholders. This may involve conducting usability studies, gathering feedback from end-users, and iteratively refining the framework based on user input.

Literature

[1]Been Kim, Thynthia Rudin, Julie Sah, “The Bayesian CASE model: a generative approach for case-based reasoning and prototype classification”, *Neural Information Processing Systems (NIPS)*, 2014.

[2]Been Kim, Elena Glassman, Brittney Johnson, Julie Shah, “iBCM: interactive Bayesian case model empowering humans via intuitive interaction”, *Computer Science and Artificial Intelligence Laboratory Technical Report*, 2015.

[18] Ariawan, Eko & Zefanya, Maria. (2018). Predictive Analysis of Employee Loyalty: A Comparative Study Using Logistic Regression Model and Artificial Neural Network.