РЕЗЮМЕТА

НА ТРУДОВЕТЕ СЛЕД ЗАЩИТА НА ДОКТОРСКА ДИСЕРТАЦИЯ

на гл.ас. д-р Антоанета Петрова Йорданова

за участие в конкурс за заемане на академична длъжност „Доцент"
в област на висше образование 4. Природни науки, Математика и Информатика,
професионално направление 4.5. Математика, по научна специалност „Математическо
моделиране и приложение на математиката"

Predicting the 305-Day Milk Yield of Holstein-Friesian Cows Depending on the Conformation Traits and Farm Using Simplified Selective Ensembles

Snezhana Gocheva-Ilieva [1,*], Antoaneta Yordanova [2] and Hristina Kulina [1]

[1] Department of Mathematical Analysis, University of Plovdiv Paisii Hilendarski, 24 Tzar Asen St., 4000 Plovdiv, Bulgaria; kulina@uni-plovdiv.bg
[2] Medical College, Trakia University, 9 Armeyska St., 6000 Stara Zagora, Bulgaria; antoaneta.yordanova@trakia-uni.bg
* Correspondence: snow@uni-plovdiv.bg

Abstract: In animal husbandry, it is of great interest to determine and control the key factors that affect the production characteristics of animals, such as milk yield. In this study, simplified selective tree-based ensembles were used for modeling and forecasting the 305-day average milk yield of Holstein-Friesian cows, depending on 12 external traits and the farm as an environmental factor. The preprocessing of the initial independent variables included their transformation into rotated principal components. The resulting dataset was divided into learning (75%) and holdout test (25%) subsamples. Initially, three diverse base models were generated using Classifiction and Regression Trees (CART) ensembles and bagging and arcing algorithms. These models were processed using the developed simplified selective algorithm based on the index of agreement. An average reduction of 30% in the number of trees of selective ensembles was obtained. Finally, by separately stacking the predictions from the non-selective and selective base models, two linear hybrid models were built. The hybrid model of the selective ensembles showed a 13.6% reduction in the test set prediction error compared to the hybrid model of the non-selective ensembles. The identified key factors determining milk yield include the farm, udder width, chest width, and stature of the animals. The proposed approach can be applied to improve the management of dairy farms.

Keywords: machine learning; rotation CART ensemble; bagging; boosting; arcing; simplified selective ensemble; linear stacked model

Stacking Machine Learning Models using Factor Analysis to Predict the Output Laser Power

Snezhana Gocheva-Ilieva
Faculty of Mathematics and Informatics
Plovdiv University Paisii Hilendarski
Plovdiv, Bulgaria
snow@uni-plovdiv.bg

Hristina Kulina
Faculty of Mathematics and Informatics
Plovdiv University Paisii Hilendarski
Plovdiv, Bulgaria
kulina@uni-plovdiv.bg

Antoaneta Yordanova
Medical College
Trakia Unversity
Stara Zagora, Bulgaria
antoaneta.yordanova@trakia-uni.bg

Abstract— Machine learning is an intensively evolving field of artificial intelligence with applications in all areas where data from observations and experiments are accumulated. In this study, for the first time, a new approach is developed to combine the predictions of several diverse machine learning models based on the stacked generalization paradigm. A new framework for stacking the models using exploratory factor analysis has been proposed. The approach is demonstrated for predicting the output laser electric power of a copper bromide vapor laser. The achieved results show the superiority of the constructed stacked model compared to the single base models.

Keywords—stacking paradigm, factor analysis, output laser power, copper bromide vapor laser, ensemble model

developed and many applied results have been obtained with the application of ML and the stacking paradigm.

A stacked ensemble neural networks models based on long short-term memory (LSTM) and convolutional-LSTM as a meta model is developed in [10] for predicting turbocharger remaining useful life using high dimensional sensor data. In [11] the authors applied various ML models, such as extreme gradient boosting (XGBoost), random forest (RF), classification and regression trees (CART), and M5 tree methods to predict energy performance in residential buildings. The same methods techniques were also used as to build the stacked models. As a result, the XGboost stacked model is identified with the highest accuracy level.

Антоанета Йорданова

ПРИЛОЖЕНИЕ НА ДЕЙТА МАЙНИНГ МЕТОДИ ЗА СТАТИСТИЧЕСКО МОДЕЛИРАНЕ

ПЛОВДИВСКО УНИВЕРСИТЕТСКО ИЗДАТЕЛСТВО

Article

Population Pharmacokinetics of Doxycycline, Administered Alone or with N-Acetylcysteine, in Chickens with Experimental *Mycoplasma gallisepticum* Infection

Tsvetelina Petkova [1], Antoaneta Yordanova [2] and Aneliya Milanova [1,*]

1   Department of Pharmacology, Animal Physiology, Biochemistry and Chemistry, Faculty of Veterinary Medicine, Trakia University, 6000 Stara Zagora, Bulgaria

2   Department of Social Medicine, Healthcare Management and Disaster Medicine, Faculty of Medicine, Trakia University, 6000 Stara Zagora, Bulgaria

*   Correspondence: akmilanova@gmail.com or aneliya.milanova@trakia-uni.bg; Tel.: +35-942-699-696

Abstract: Mycoplasmosis is a bacterial infection that significantly affects poultry production, and it is often controlled with antibiotics, including doxycycline. The conducted study aimed to determine population pharmacokinetic (PopPk) parameters of doxycycline in healthy (n = 12) and in *Mycoplasma gallisepticum*-challenged (n = 20) chickens after its oral administration via drinking water at the registered dose rate of 20 mg/kg b.w./24 h for five days, without or with co-administration of N-acetylcysteine (NAC, a dose of 100 mg/kg b.w./24 h) via the feed. Doxycycline concentrations in plasma were analyzed with the LC-MS/MS method. The values of tvV/F and tvk_e were $4.73 \text{ L} \times \text{kg}^{-1}$ and 0.154 h^{-1}, respectively, and they showed low BSV. A high BSV of 93.17% was calculated for the value of tlag of 0.8 h, which reflects the inter-individual differences in the water consumption. PTA was computed after Monte Carlo simulation with the registered dose for doxycycline. The target of $\%fT > \text{MIC} \geq 80\%$ and 100% can be achieved in 90% of the broiler population, after a correction for protein binding, for bacteria with $\text{MIC} \leq 0.5 \text{ mg} \times \text{L}^{-1}$ and $0.25 \text{ mg} \times \text{L}^{-1}$, respectively. The applied PopPk model did not reveal significant effect of *M. gallisepticum* infection and co-administration of NAC on pharmacokinetic parameters of doxycycline.

Keywords: doxycycline; mycoplasmosis; N-acetylcysteine; population pharmacokinetics; poultry

Dependence of Body Mass Index on Some Dietary Habits: An Application of Classification and Regression Tree

Magdalena Platikanova [1], Antoaneta Yordanova [2], Petya Hristova [1]

1. Department of Hygiene, Epidemiology, Microbiology, Parasitology and Infectious Diseases, Medical Faculty, Trakia University, Stara Zagora 6000, Bulgaria
2. Medical College, Trakia University, Stara Zagora 6000, Bulgaria

Corresponding Author: Email: mplatikanova@abv.bg

Abstract

Background: The purpose of this study was to determine the influence of some eating habits on body mass index (BMI) using a regression model created via the classification and regression tree method (CART).

Methods: The study was conducted using a questionnaire specially developed for the study, evaluated for reliability and validity. In addition to demographics (age and sex), the questions concern the timing of the meals and the type of food consumed. The data contains records for 533 people (322 women and 211 men) aged 18 to 65 years. The survey was conducted in the period 2019-2021 in Stara Zagora, Bulgaria. Data were processed using descriptive statistics, and regression and classification data mining method CART.

Results: A CART model with a dependent variable BMI and predictors Sex, Age, Breakfast type, Breakfast time, Lunchtime, Lunch type, Dinner time, Dinner type have been created. The obtained model is statistically significant at a significance level of $P<0.0001$ and a coefficient of determination $R^2 = 0.495$. The normalized importance of the factors that affect the BMI is as follows: Sex (100%), Age (61.4%), Lunch type (26. 0%), Lunchtime (18.8%), Dinner time (13.9%), and Breakfast type (13.2%). Women have a lower BMI than men. BMI increases with age.

Conclusion: The CART method allows to make a classification by the predictors used and gives opportunities for a more in-depth analysis of the reasons for the increase in BMI. The level of influence of diet and eating habits (type of food, time of consumption) on BMI was determined.

Keywords: Body mass index (BMI); Dietary habits; Classification and regression tree (CART method)

Criteria for selection of statistical data processing software

Lina Yordanova[1], Gabriela Kiryakova[1], Petya Veleva[1], Nadezhda Angelova[1], Antoaneta Yordanova[2]

[1]Trakia University, Stara Zagora, Bulgaria
[2]University of Plovdiv Paisii Hilendarski, 24 Tsar Asen Str., 4000 Plovdiv, Bulgaria

e-mail: lina.yordanova@trakia-uni.bg, nadezhda.angelova@trakia-uni.bg

Abstract. There are many statistical software packages that automate the activities related to statistical data processing and visualization of results. Often the choice of a statistical package by consumers is a difficult process and it is necessary to compare different products in order to highlight the most appropriate. The selection of one or another statistical package should be based on clear criteria and be made in accordance with the objectives set by the researchers. In the present work, the criteria for comparing statistical packages are systematized into the three groups and applied to four packages.

Since 2002

OPEN ACCESS

Early Seizures during Stroke

Christiyan Naydenov[1]*, Velina Mancheva[1], Lachezar Manchev[1], Antoaneta Yordanova[2]

[1]Department of Neurology, Medical Faculty, Trakia University, Stara Zagora City, Bulgaria; [2]Department of Health Care, Medical college, Trakia University, Stara Zagora city, Bulgaria

Abstract

BACKGROUND: The connections between seizures and stroke, their characteristics, and features are barely developed and also very important for the outcome. The diagnostic problem leads to ineffective treatment due to difficult selection of patients who are subject to prevention with antiepileptic drugs (AEDs) on one hand. On the other hand, it is impossible to exam every stroke patient by EEG. We need an algorithm for screening indicated patients and conducting the EEG. After that, we can include properly AED as a prevention. Their low frequency makes them time consuming to study.

AIM: The aim of the study was to conduct an epidemiological study of early epileptic seizures at the acute stroke phase and to derive principles for screening, diagnosis, and behavior for prevention.

METHODS: To achieve the goal, we have researched retrospective patients, totally amounting to 656.

RESULTS: The factors identified so far in the genesis of seizures such as age, type, and location of cerebrovascular accident and proximity to the cortex do not act alone, but in combination with undiscovered ones.

CONCLUSIONS: There are no clear criteria to outline the rules for the AED prophylactic in patients with cerebrovascular disease (CVD). The most important indicator is the systematic assessment of the risk of seizures in the course of the disease. Patients at high risk of triggering seizures and developing epilepsy in CVDs are suitable for EEG examinations with a view to timely diagnosis and treatment.

Since 2002

Methodology for EEG and Reference Values of the Software Analysis

Christiyan Naydenov[1]*, Antoaneta Yordanova[2], Velina Mancheva[1]

[1]Department of Neurology, Medical Faculty, Trakia University, Stara Zagora, Bulgaria; [2]Medical College at the Trakia University, Stara Zagora, Bulgaria

Abstract

BACKGROUND: EEG is a way of graphically recording the electrical potentials of the brain. The main parameters of the EEG are related to the ratios between the individual frequency components, their amplitude and characteristic waveforms.

AIM: The aim of the study was to develop and describe a consistent and detailed methodology for the technical conduct of an EEG study, as well as to find the reference values of some of the most frequently derived average values of parameters from the software analysis of modern EEG equipment. TOOLKIT: The EEG office of the Medical Faculty at Trakia University is equipped with a multifunctional 31-channel digital EEG/EP device with a sampling frequency of 1000 Hz.

METHODOLOGY: Unification of the technical implementation is necessary so that the results can be compared with the maximum cleared probability of statistical error due to controllable factors. The outlined sequential steps can serve any technical contributor.

RESULTS: Electroencephalographic study of 100 clinically sound participants and determination of the reference values of the indicators. We have created a table with the mean values of the EEG Software Analysis.

CONCLUSION: The frequency composition of the EEG signal includes four types of waves or rhythms: delta/0.5−3.5Hz/, theta/4−7.5Hz/, alpha/8−12Hz/,and beta/13−30Hz/. The amplitude of the EEG under physiological conditions ranges from 15 microvolts/low-amplitude/to 150 microvolts/high-amplitude/. Average limits are between 30 and 80 microvolts.

Since 2002

Determination of Factors Affecting Vitamin D Levels in Women at Risk Using Classification and Regression Tree Analysis

Antoaneta Yordanova[1]*, Anna Tolekova[2], Pavlina Teneva[2], Ivelina Dobreva[2], Katya Kichukova[2]

[1]Department of Social Medicine, Healthcare Management and Disaster Medicine, Trakia University, Stara Zagora, Bulgaria; [2]Department of Health Care, Medical College, Trakia University, Stara Zagora, Bulgaria

Abstract

BACKGROUND: The influence of the factors on Vitamin D as a health indicator in premenopausal and menopausal women is a significant subject to be investigated.

AIM: The study uses the potential of classification and regression trees (CART) as a data mining method for medical type samples.

METHODS: The data set is built by records of 84 indoor working women at the age of 45 to 67 years from five Bulgarian companies. The data are obtained through laboratory tests of serum concentrations of 25-OH-Vitamin D and a questionnaire, created for the study. Statistical data processing is made by descriptive statistics and the CART method.

RESULTS: The results show Vitamin D deficiency in 69% of the studied women at risk from Stara Zagora. For the target variable – Vitamin D (the quantity of 25-OH-Vitamin D), a regression CART tree was built. The calculated percentages of normalized importance for each independent variable reveal that the most important predictors, affecting Vitamin D, are body mass index (100%), alcohol (84.2%), education (70.3%), coffee (70.2%), Ca_Vit D (69.8%), and sports frequency (42.4%), while the other variables have much less importance.

CONCLUSION: The application of the CART method makes it possible to study the distribution and importance of the factors influencing the state of vitamin D. The presence of such a high percentage of women at risk requires a comprehensive approach, including educational programs and strict application of guidelines for vitamin D supplementation to prevent the effects of hypovitaminosis.

Лина Йорданова
Габриела Кирякова
Петя Велева
Надежда Ангелова
Антоанета Йорданова

Статистически софтуер за обработка на експериментални данни

Стара Загора
2019 г.

Предговор

Провеждането на научни изследвания е свързано със събиране, подготовка и обработка на експериментални данни. Изборът на подходящи статистически методи за обработка на данните от проучвания има изключително важно значение за постигането на целите на научното изследване.

Целите на настоящата монография включват:
• създаване на система от критерии за сравнение на статистически софтуерни пакети;
• проучване и сравнение на локални софтуерни пакети, Web услуги и средства в реално време за обработка на експериментални данни;
• представяне възможностите на избрани статистически софтуерни средства чрез реална обработка на примерни данни.

В монографията е реализирана идеята за интегриране на съвременни технологии към печатните материали, по-конкретно Добавена реалност (Augmented Reality) и Облачен изчислителен модел (Cloud computing). Разработено е мобилно приложение BookStatistica, чрез което читателите имат достъп до допълнителни материали (видео, аудио и други формати), които онагледяват разглежданите статистически обработки и анализи и допринасят за по-лесното им разбиране.

В **Глава 1. Обзор на статистическите ресурси и софтуерните пакети** е направен преглед на различни видове ресурси по статистика – виртуални библиотеки, университетски сайтове за обучение, мобилни приложения, видео ресурси, средства за визуализация, както и възможностите за онлайн статистическа обработка.

В **Глава 2. Въведение в статистическите методи** са разгледани същността, предметът и задачите на статистическите методи. Обърнато е внимание на важността на процесите на планиране и анализ на експеримента, моделирането на процесите и явленията. Дефинирани са основни понятия като генерална съвкупност, извадка, видове променливи, скали на измерване, начините на представяне на данните чрез статистически или вариационен ред или чрез честотно групиране. Представени са различни алгоритми за избор на статистически метод, както и Web-базирани системи, подпомагащи избора.

В **Глава 3. Сравнение на софтуерни средства за статистическа обработка на данни** са дефинирани критерии, на базата на които избрани софтуерни продукти – Statistica, SPSS, XLSTAT и R Project са оценени и сравнени. Представени са основните понятия и термини и особеностите на избраните пакети.

В **Глава 4. Описателни статистически методи** с избраните статистически продукти са представени Дескриптивната статистика като един от най-често използваните инструменти за обобщено представя-

не на дадено множество от данни и основните видове графики, които представят нагледно данните и спомагат за по-лесното им и бързо възприемане от потребителите. Представени са различни видове разпределения и методи за проверка нормалността на разпределение. Реализацията на разгледаните процедури е извършена чрез избраните статистически продукти, направена е интерпретация и сравнение на получените резултати.

В **Глава 5. Инферентни методи** са разгледани различни варианти при проверка на статистически хипотези, дисперсионен, корелационен и регресионен анализ. Разгледаните процедури са реализирани чрез избраните статистически продукти, като е предложено тълкуване и е направено сравнение на резултатите.