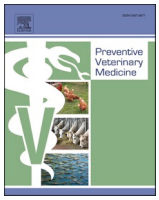
Preventive Veterinary Medicine 230 (2024) 106291

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed



Anomaly detection in the veterinary antibiotic prescription surveillance system (IS ABV)

Guy-Alain Schnidrig [a,c,*], Anaïs Léger [b], Heinzpeter Schwermer [b], Rebecca Furtado Jost [b], Dagmar Heim [b], Gertraud Schüpbach-Regula [a]

[a] Veterinary Public Health Institute, Vetsuisse, University of Bern, Schwarzenburgstrasse 161, Liebefeld 3097, Switzerland
[b] Federal Food Safety and Veterinary Office (FSVO), Schwarzenburgstrasse 155, Bern 3003, Switzerland
[c] Graduate School of Cellular and Biomedical Sciences, University of Bern, Hochschulstrasse 4, Bern, Switzerland

ARTICLE INFO

Keywords:
Antibiotic Use
Cattle
IS ABV
Machine Learning
Surveillance

ABSTRACT

Antibiotic resistance is one of the major concerns in veterinary and human medicine and poses a considerable threat to both human and animal health. It has been shown that over- or misuse of antibiotics is one of the primary drivers of antibiotic resistance. To develop the surveillance of antibiotic use, Switzerland introduced the "Informationssystem Antibiotika in der Veterinärmedizin" (IS ABV) in 2019, mandating electronic registration of antibiotic prescriptions by all veterinarians in Switzerland. However, initial data analysis revealed a considerable amount of implausible data entries, potentially compromising data quality and reliability. These anomalies may be caused by input errors, inaccuracies, incorrect or aberrant master data or data transmission and make analysis impossible. To address this issue efficiently, we propose a two-stage anomaly detection framework utilizing machine learning algorithms. In this study, our primary focus was on cattle treatments with either single or group therapy, as they were the species with the highest prescription volume. However, not all outliers are necessarily incorrect; some may be legitimate but unusual antibiotic treatments. Thus, expert review plays a crucial role in distinguishing outliers, that are correct from actual errors. Initially, relevant prescription variables were extracted and pre-processed with a custom-built scaler. A set of unsupervised algorithms calculated the probability of each data point and identified the most likely outliers. In collaboration with experts, we annotated anomalies and established anomaly thresholds for each production type and active substance. These expert-annotated labels were then used to fine-tune the final supervised classification algorithms. With this methodology, we identified 22,816 anomalies from a total of 1,994,170 prescriptions in cattle (1.1 %). Cattle with no further specified production type had the most (2 %) anomalies with 7758 out of 379,995. The anomalies were consistently identified and comprised prescriptions with too high and too low dosages. Random Forest achieved a ROC-AUC score of 0.994, (95 % CI: 0.992, 0.995) and a F1-Score of 0.962 (95 % CI: 0.958, 0.966) for single treatments. The versatility of this framework allows its adaptation to other species within IS ABV and potentially to other prescription-based surveillance systems. If applied regularly to uploaded prescriptions, it should reduce input errors over time, improving the validity of the data in the long term.

1. Introduction

One of the main issues in veterinary and human medicine is antibiotic resistance, which poses a considerable threat to both human and animal health (Prestinaci et al., 2015; Laxminarayan et al., 2013). Mechanisms for antibiotic resistance have existed and evolved over a long period of time, enabling bacteria to rapidly adapt to external selective pressures (Giedraitienė et al., 2011; Perry et al., 2016). It has been shown that over- or misuse of antibiotics creates a selective environment, and is one of the primary drivers of antibiotic resistance (Bronzwaer et al., 2002; Goossens et al., 2005; Van De Sande-Bruinsma et al., 2008; Caneschi et al., 2023).

Surveillance of antimicrobial use (AMU) and resistance is essential to guide appropriate interventions. Several European countries have implemented monitoring systems to investigate the use of antibiotics in veterinary medicine (Sanders et al., 2020; European Medicines Agency,

* Corresponding author at: Veterinary Public Health Institute, Vetsuisse, University of Bern, Schwarzenburgstrasse 161, Liebefeld 3097, Switzerland
E-mail address: schnidrig.guy@gmail.com (G.-A. Schnidrig).

https://doi.org/10.1016/j.prevetmed.2024.106291
Received 15 March 2024; Received in revised form 12 July 2024; Accepted 14 July 2024
Available online 19 July 2024

2013). Previously, AMU in Switzerland was merely monitored through the Veterinary Antibiotics sales figures (FSVO, 2020a). Although the sales figures for antibiotics in Switzerland showed an overall decreasing trend, it was not possible to identify the species or specific animal sectors for which the antibiotics were used. The development of AMR is not mainly determined by the amount of active substances administered to a population but also by the number of treated animals and other factors (Caneschi et al., 2023).

To improve the surveillance of AMU in Switzerland, the Federal Food Safety and Veterinary Office (FSVO) has introduced the «Informationssystem Antibiotika in der Veterinärmedizin» (IS ABV) (FSVO, 2020b). Since October 2019, every veterinarian in Switzerland must register all antibiotic prescriptions electronically in a centralized database. prescription is both dispensed to the animal owner and used by the veterinarian.

Although the implementation of the IS ABV was a significant advancement, it has revealed several challenges and issues. In 2020 initial analyses of the data showed that there was a considerable amount of implausible data entries leading to enormous amounts of active substance. This indicated that there were systemic or human errors in data entry and transmission process, compromising the reliability and accuracy of the data. With 1,768,959 prescriptions across all species from 2019 until the end of 2020, the FSVO was only able to identify the most anomalies, which were characterized by obvious errors resulting in a disproportionately high quantity of prescribed antibiotic substance per animal (FSVO, 2021). As such anomalies could be due to erroneous inputs distributed over several data fields, we were interested in a flexible data-driven method over fixed rules.

To address this problem, we propose a two-stage detection framework that utilizes a combination of several machine learning algorithms and is designed to identify potential anomalies.

Unsupervised learning uses algorithms to identify patterns or structures in datasets without any external guidance or labeled information (Amruthnath and Gupta, 2018). In contrast to supervised learning, where the algorithm is provided with a first labeled dataset to predict the outcomes of new/the next datasets, unsupervised learning allows the model to explore the underlying data structure without being given any specific task or instruction (Sarker, 2021). One major disadvantage of unsupervised learning is that there are not many adequate ways to evaluate performance when limited labels are available. However, an initial unsupervised approach for anomaly or outlier detection could still be a compelling method in identifying and flagging outliers given the scarcity of labeled data. Traditionally, outliers or anomalies are data points that differ significantly from the majority of the data and can indicate errors, different underlying mechanisms, or uncommon behavior (Aggarwal, 2017). In the context of the IS ABV system, we are interested in identifying outliers that are indicative of input errors in a variety of input fields, application inaccuracies or instances of severe over- or under-dosage respectively. Nevertheless, it is important to note that not all identified outliers result from errors. Certain outliers may be genuine data points that represent legitimate but rare instances of an antibiotic treatment. Hence, it is crucial to have an expert review of any identified outliers to determine whether they are genuine data points or not.

The objective of this study is to present a two-stage framework that aims to support and enhance human expertize in identifying potential outliers rather than replacing it. In the discovery stage, unsupervised algorithms initially computed the probability of each data point to be an outlier and determine the most likely outliers. Together with experts we then annotated anomalies and established anomaly thresholds per active substance and production type. In the detection stage we employed supervised methods to refine the anomaly detection system which ultimately led to a more robust and reliable outlier detection.

2. Methods

2.1. Data collection

In Switzerland, veterinarians have a choice between two methods for submitting their prescriptions to IS ABV. They can either manually enter the prescriptions through a website (variant 2), or they can connect their practice management software to the ISABV server-interface (variant 1). Variant 1 is substantially more common and accounts for 94.6 % of all prescriptions over the past three years. IS ABV data are available from the FSVO in an anonymized form for research purposes upon request, which we made for the first time in an electronic form on 15 April 2022 (see data availability).

We aimed to capture a comprehensive representation of outliers and therefore included prescriptions from several years. We extracted the data on January 17th, 2023, which contained prescriptions from January 1st, 2020, up to January 16th, 2023. Within the specified timeframe, every prescription issued by veterinary practices and clinics registered in IS ABV was included.

In this study we concentrated on the farm animal species that received the highest number of prescriptions, namely cattle. Our focus was on the most prevalent and national important animal livestock species (FSVO, 2020b). A total of 1,994,170 unique cattle prescriptions were included which involved 8 distinct cattle production types (Table 1). The selected data accounts for 73.8 % of all farm animal prescriptions administered in the past three years.

For livestock, in IS ABV there are four possible prescription types: single treatment (ST), oral group treatment (GT) and dispensing on stock (DoS). DoS is a specific practice in Switzerland where veterinarians can, under certain conditions, provide antibiotics to farmers, who can then apply it them themselves whilst prophylactic use is forbidden. It is important to note that DoS prescriptions, unlike the three others, do not include information about the production type and number of animals treated, and were therefore excluded. GT are prescriptions that are exclusively done manually (variant 2) over the web interface and contain orally applied antibiotic treatments for groups of 10 or more animals. ST are all other antibiotic prescriptions that are not defined as DoS or GT. Up until September 2021, ST could only contain one animal per prescription. Since then, ST can also be used for multiple animals. Furthermore, a single ST prescription can contain multiple preparations and one preparation can contain multiple active substances. The production types with the highest number of prescriptions (Table 1) were dairy cows (1,138,935; 57.1 %) and cattle with no production type specified (379,995; 19.1 %).

During the initial data pre-processing stage, we identified and removed entries with null values from the prescription dataset (Fig. 1). Null values in prescriptions refer to instances where the recorded amount of antibiotics given, or the number of treated animals is zero. We also excluded some prescriptions with rare AB which should not be used

Table 1
Prescriptions and anomalies per category of use.

Category of use	Prescriptions (%)	Prescriptions with Anomalies	Percent of prescriptions containing anomalies
Dairy cow	1,138,935 (57.11)	10,534	0.92
Cattle with no production type	379,995 (19.06)	7758	2.04
Rearing calves	155,693 (7.81)	1279	0.82
Veal calves	119,973 (6.02)	1240	1.03
Suckler cow	74,629 (3.74)	755	1.01
Rearing cattle	55782 (2.80)	572	1.03
Suckler calf	51,069 (2.56)	442	0.87
Beef cattle	18,094 (0.91)	236	1.31
Total	1,994,170 (100)	22,816	1.14

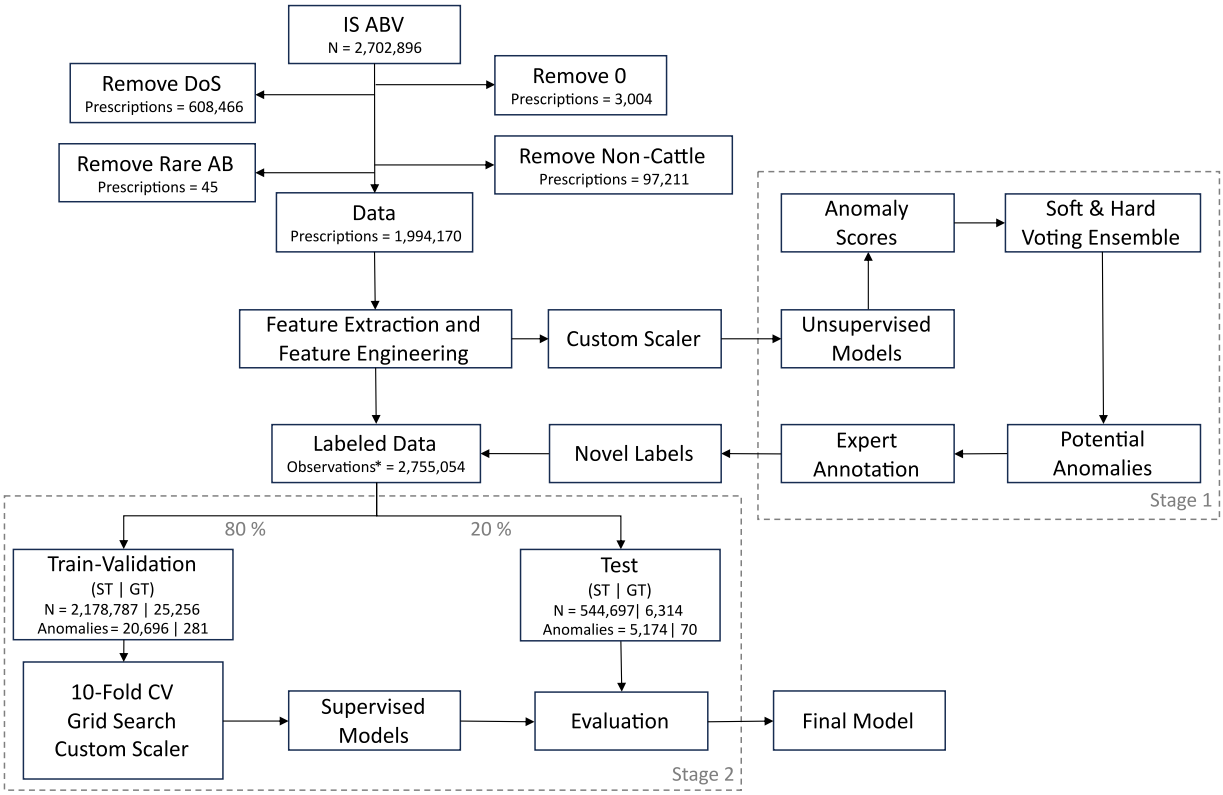


Fig. 1. Flowchart of the data extraction and the two-stage anomaly detection process. *Number of observations is higher than number of prescriptions because one prescription can contain more than one preparation with more than one active substance.

on cattle and/or have less than 5 prescriptions within the past 5 years.

2.2. Feature engineering and data preparation

To identify anomalies, we established the variable *given amount per animal and day* (GAAD). This variable considers the quantity (mg) of antibiotic active ingredients given per animal and the duration of the treatment (days). By normalizing the amount of antibiotics based on the number of animals and treatment duration we were able to compare different prescriptions with each other.

We applied the natural logarithmic transformation of $(x + 1)$ to the GAAD. The dataset was then separated by prescription type (ST and GT). GAAD was analyzed in four different groups by production type, preparation, and active substance. The first groups GAAD by production type and preparation id (product identification key). The second groups GAAD per preparation id. The third groups GAAD by production type and active substance, and the fourth groups GAAD per active substance. GAAD was normalized per group using z-transformation. With a custom scaler multiple z-transformations (Lynn, 1986) were conducted to create four variables which distinguish the log transformed GAAD (X_i) per group (g) on different levels:

$$Z_g = \frac{X_i - \mu_g}{\sigma_g}$$

The custom scaler is based on the *z-score* from scipy.stats (Virtanen et al., 2020) and the *BaseEstimator* as well as the *TransformerMixin* from scikit-learn (Pedregosa et al., 2011).

Evaluation and analysis were conducted using Python (van Rossum, 1995; version 3.9.18) with Jupyter Notebook (Kluyver et al., 2016; version 6.5.4) and R (version 4.2.1) with R Studio (RStudio, 2015; version 2023.09.0). All computations were done with an AMD Ryzen 7 3700×8-Core Processor, 32 GB of RAM, and an NVIDIA GeForce GTX 3080 graphics card.

2.3. First stage analyses

We used the python library called PyOD (Zhao et al., 2019) which offers a convenient wrapper for the most common unsupervised anomaly detection algorithms and is based on scikit-learn (Pedregosa et al., 2011). For each of the two data sets (ST and GT), five unsupervised algorithms were implemented.

Empirical Cumulative Outlier Detection (ECOD) uses empirical cumulative distribution functions to identify anomalies (Li et al., 2022). ECOD estimates the distribution of the input data by computing the empirical cumulative distribution per dimension and then estimates tail probabilities for each data point and computes an outlier score.

Gaussian Mixture Model (GMM) is a probabilistic model that assumes that the data is generated from a mixture of Gaussian distributions (Aggarwal, 2017). To detect anomalies, it starts by randomly selecting subsets of the data, then for each random subspace it estimates the probability density using a gaussian mixture model. Geometric averaging is performed to get an overall probability density from which then an outlier score is calculated. Prescriptions which have a low probability density compared to the overall distribution are flagged as anomalies.

Histogram-Based Outlier Score (HBOS) is an anomaly detection algorithm that utilizes histograms to calculate outlier scores (Goldstein and Dengel, 2012). The algorithm partitions the data into bins for each dimension, then it constructs a histogram for each dimension which counts the number of data points that fall within each bin. The probability that a prescription occurs in each histogram is then used to compute an outlier score.

Isolation Forest (IForest) identifies anomalies directly by applying a tree structure to isolate every data point (Morales et al., 2020). The algorithm randomly selects a subset of data points from the dataset and creates a tree structure called isolation tree. Each isolation tree is constructed by recursively partitioning the data points based on random

attribute splits until all data points are isolated. Based on the average path length across all isolation trees it then calculates an outlier score.

Principal Component Analysis (PCA) is an established statistical technique used for reducing the dimensionality of data (Daga et al., 2020). PCA can also be used for detecting outliers by utilizing a robust principal component classifier (Shyu et al., 2003). It constructs an intrusion predictive model using the major and minor principal components. The implemented version assumes that the data follows a multivariate Gaussian distribution and that outliers are defined as points with low probability.

2.3.1. Parameter tuning first stage

For most unsupervised anomaly detection methods, it is necessary to manually set a contamination factor which determines the percentage of points in the data that are considered anomalous. The contamination parameter per model and species was based on previously discovered anomalies and known prescription issues in certain production types and preparations. Anomalies that were already tagged consisted only of a few instances with very high dosages. To detect known outliers as well as identify novel prescriptions with insignificant dosages, we iteratively adjusted the initial parameter settings until convergence no longer improved, or satisfactory results were achieved. For the first stage we only included the four GAAD variables as describe in Data preparation. The resulting anomaly scores were then combined and assessed using a soft (probability) and hard (binary) voting ensemble. This resulted in a table which ranked the anomalies from most probable to least probable. In collaboration with experts from the FSVO, we evaluated the prescriptions and established specific GAAD based thresholds for each production type and active substance. Together with manual adjustments and corrections per preparation and production type, everything that exceeded these thresholds of the soft and hard votes was a novel anomaly and served as additional labels for the second stage. To summarize, in this study an anomaly is defined as any observation that exceeds the expert guided thresholds and thus has a GAAD that is outside the normally expected range.

2.4. Second stage analyses

By combining established and newly labeled prescriptions with anomalies, we were able to apply two supervised models:

Random Forest is an ensemble learning method that combines the predictions of multiple decision trees into an unified outcome (Ho, 1995). It uses both bagging and variable randomness to create an uncorrelated forest of decision trees. Random Forest is designed to address the problem of overfitting and reduces the variance in decision trees. We implemented the Random Forest classifier from scikit learn (Pedregosa et al., 2011).

XGBoost is also an ensemble learning method and stands for Extreme Gradient Boosting (Chen and Guestrin, 2016). It combines multiple weaker models and applies a gradient boosting technique, which sequentially builds a predictive model by fitting each new model to the residual errors of the previous predictions.

2.4.1. Parameter tuning second stage

To achieve an impartial estimate, we split the labeled data into 80 % training-validation set and a 20 % holdout set (Fig. 1). To ensure an equal distribution of classes (inlier and outlier) in both sets, we implemented a stratified split based on the labeled anomalies. In addition to our five initial unsupervised models, we implemented two common supervised models, Random Forest and XGBoost. We also reintroduced active substance and production type as “one hot encoded” categorical variables (Supplement Table 3). To fine-tune the hyperparameters of the supervised algorithms, an initial random and then an extensive grid search was conducted on the training-validation set using stratified 10-fold cross-validation (Supplement Table 5; Supplement Table 6). Missing values generated by the z-transformation within cross-validation and training steps were imputed with 0. We optimized the hyperparameters from the initial unsupervised models by assessing the performance of each hyperparameter combination in every fold and selecting the configuration that yielded the best F1-Score. To avoid introducing data leakage, the unsupervised models were re-trained, re-scaled and had their hyperparameters re-adjusted on the train-validation set.

2.5. Model evaluation

To compare the models’ performances on the holdout set we employed the following techniques: ROC-AUC (Bradley, 1997), Precision, Recall, F1-Score, MCC (Baldi et al., 2000) and AP (Zhang and Zhang, 2009). The evaluation metrics are described in more detail in the supplement. Confidence intervals (95 %) for each metric were generated using bootstrap resampling (n = 1000) on the holdout set.

2.6. Post-hoc analysis

Post-hoc analysis was performed to assess the importance of individual variables for the model prediction using SHAP. SHAP has a theoretical foundation in coalitional game theory and can provide contribution explanations and analyze the model’s output on a global and a local scale (Lundberg and Lee, 2017). Here SHAP is used to quantify the relative importance of variables contributing towards anomaly. An average higher (positive) SHAP indicates a higher impact on the model’s prediction of an anomaly. SHAP values and plots were produced with the SHAP library (Lundberg and Lee, 2017) on the holdout set using *shap.TreeExplainer*. SHAP utilizes a game-theoretic approach to model explanation, assessing the individual contribution of each variable to the overall prediction for a single observation.

Model calibration is an essential practice that ensures predictive models’ reliability and interpretability (Van Calster et al., 2019). It adjusts the output probabilities of the supervised models to better reflect the true underlying probabilities of the anomalies. Model calibration was implemented from scikit-learn (Pedregosa et al., 2011) using Platt scaling (Platt, 2000) and isotonic regression (Niculescu-Mizil and Caruana, 2005). The Brier Score (Winkler, 1994) and Log Loss (Vovk, 2015) assess the agreement between predicted probabilities and actual classification results, with a lower score indicating better calibration. Both metrics are used to evaluate the performance of a classification model.

3. Results

3.1. First stage - discovery

In the first stage and with the previous labels we discovered and classified 22,816 prescriptions as anomalies. An example of the algorithms’ abilities to distinguish anomalies is shown in a two-dimensional representation in Fig. 2. The remaining contour plots can be found in Supplement Fig. 1 and Supplement Fig. 2.

Out of a total of 1,994,170 cattle prescriptions, 1.14 % of them had an anomaly. 22,599 anomalies were found in prescriptions with single treatment and 217 anomalies were found in prescriptions with group treatments. The production type dairy cows exhibited most anomalies with 10,534 out of 1,138,935 prescriptions (0.92 %) (Table 1). The lowest number of anomalies were found in beef cattle with 236 out of 18,094 (1.31 %) prescriptions. Cattle with no production type had the highest proportion of anomalies with 7758 anomalies out of 379,995 (2.04 %) prescriptions. The lowest relative percentage of anomalies were found in rearing calves with 572 out of 55782 (0.82 %) prescriptions. Suckler calves had the second lowest proportion of anomalies with 755 out of 74,629 (0.87 %) prescriptions.

The active substance class with most anomalies (Supplement Table 1) was penicillin (11,093) followed by aminoglycosides (5560). Relatively, prescriptions containing macrolides had the highest

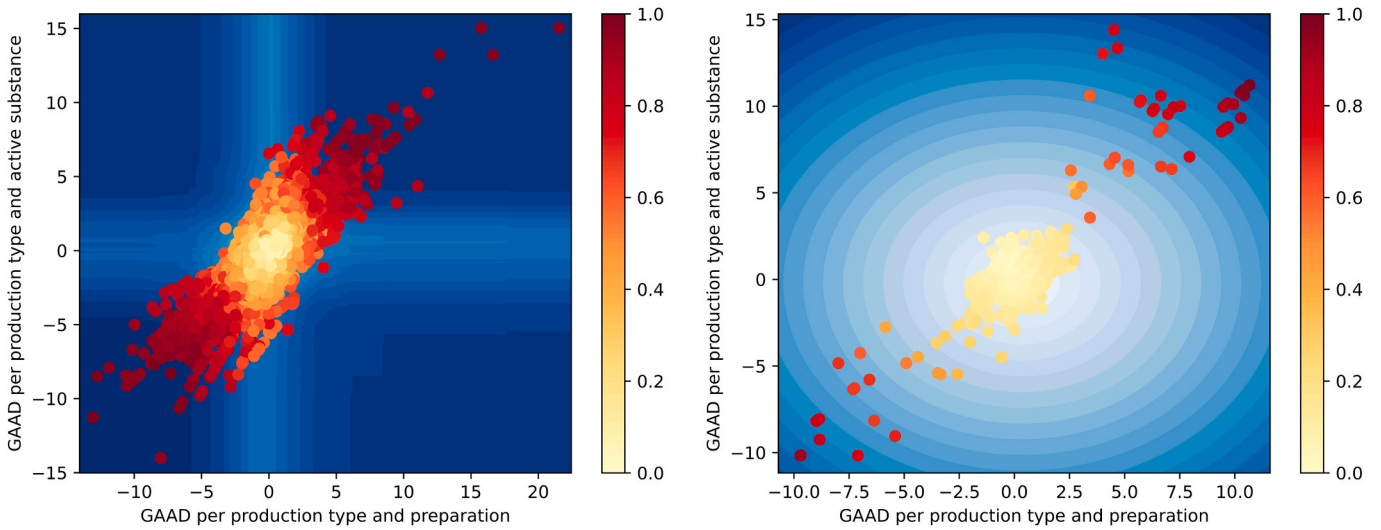


Fig. 2. Contour plots visualize the anomaly decision across two dimensions. The contour lines on the plot represent decision boundaries of the respective algorithms; IForest for the single treatments (ST) on the left and PCA for the group treatments (GT) on the right. The x-axis represents normalized GAAD (Given Amount per Animal and Day) grouped by production type and preparation. The y-axis represents the normalized GAAD grouped by production type and active substance. The color gradient represents the normalized anomaly score of each point. Darker color indicates a greater likelihood of being an outlier.

percentage of anomalies with 799 out of 45,008 prescriptions (1.78 %). The lowest number of anomalies with 5 out of 795 (0.63 %) was found in Polymyxins which were also the active substance class with lowest percentage.

The preparation categories with the highest number of anomalies (Supplement Table 2) were injection preparations (13,830) and udder injectors (4255). Relatively, the most anomalies (2.5 %) were found in tablets, capsules and bolus followed by premix for medicated feeding with 2.35 %. The lowest percentage of anomalies were found in dry cow injectors (0.56 %) and in udder injectors for lactation (0.97 %).

3.2. Second stage - detection

In the second stage we test the models' ability to detect anomalies on 20 % of unseen data, which serves as a proxy for future prescriptions. For single treatment in cattle ST Random Forest achieved the highest ROC-AUC score of 0.994 (95 % CI: 0.992, 0.995) (Table 2). The unsupervised method with ST GMM also achieved the same score of 0.994 (95 % CI: 0.992, 0.995) (Supplement Table 4). The lowest ROC-AUC score of 0.939 (95 % CI: 0.933, 0.992) was obtained by ST HBOS (Supplement Table 4). ROC-AUC scores lack variance (0.993–0.999) across all GT models expect for HBOS with a score of 0.964 (95 % CI: 0.933, 0.992).

The highest F1-Score, so the best trade-off between recall and precision, in single treatments in cattle was achieved by ST Random Forest with 0.962 (95 % CI: 0.958, 0.966). ST HBOS and ST ECOD obtained the lowest F1 scores of 0.848 (95 % CI: 0.840, 0.856) and 0.856 (95 % CI: 0.849, 0.863) (Supplement Table 4). The best F1-Score in group treatments for cattle was obtained by GT XGBoost with a score of 0.993 (95 % CI: 0.977, 1.000). The comparatively lowest trade-off between the precision and recall was achieved by GT ECOD and GT HBOS with an F1-Score of 0.966 (95 % CI: 0.934, 0.993) and 0.956 (95 % CI: 0.917, 0.986). We did not observe any major deviations between MCC and F1-Scores across both treatments and all models (Supplement Table 4).

3.2.1. Precision-recall curve

The precision-recall plot provides a representation of the trade-off between precision and recall for each model. As shown in Fig. 3, the x-axis signifies recall and the y-axis denotes precision across different thresholds. Generally, all models in cattle ST show a high precision and recall trade-off. The best performing model is ST Random Forest with an average precision (AP) of 0.997. The lowest AP was obtained by ST HBOS with 0.921. All curves for cattle group treatments show a nearly perfect match and only GT ECOD obtained a slightly different AP of 0.996.

3.2.2. Post-hoc analysis

The mean absolute SHAP value measures the average expected impact of each variable on the model output, considering all possible combinations of other variable. The degree to which each variable contributes towards the model's prediction of an anomaly is shown in Fig. 4 and Fig. 5. For ST Random Forest all four GAAD variables were the most influential variables. Independent of group, purple data points (which represent GAAD values near 0) in both beeswarm plots (Fig. 4 and Fig. 5) have all low SHAP values. For single treatments Penicillin Procaine and Cefalexin were the most impactful active substances. The most influential production type for ST Random Forest was cattle with no production type. For GT XGBoost Doxycycline and Amoxicillin were the most influential variables. The groups of GAAD ranked in 3rd to 6th in their mean absolute SHAP value. Veal calves were the most important production type in GT XGBoost.

3.2.3. Model calibration

In Supplement Fig. 3 and Supplement Table 7 it is shown that

Table 2
Predictive performance on the unseen holdout set of the supervised models divided into single treatment (ST) and group treatments (GT). 95 % confidence intervals are shown in parentheses.

Model	ROC-AUC (95 % CI)	Precision (95 % CI)	Recall (95 % CI)	F1-Score (95 % CI)	MCC (95 % CI)
ST RandomForest	0.994 (0.992, 0.995)	0.938 (0.932, 0.944)	0.988 (0.985, 0.991)	0.962 (0.958, 0.966)	0.962 (0.958, 0.966)
ST XGBoost	0.991 (0.990, 0.993)	0.902 (0.894, 0.909)	0.984 (0.980, 0.987)	0.941 (0.936, 0.945)	0.941 (0.937, 0.946)
GT RandomForest	1.000 (1.000, 1.000)	0.972 (0.927, 1.000)	1.000 (1.000, 1.000)	0.986 (0.962, 1.000)	0.986 (0.962, 1.000)
GT XGBoost	1.000 (1.000, 1.000)	0.986 (0.955, 1.000)	1.000 (1.000, 1.000)	0.993 (0.977, 1.000)	0.993 (0.977, 1.000)

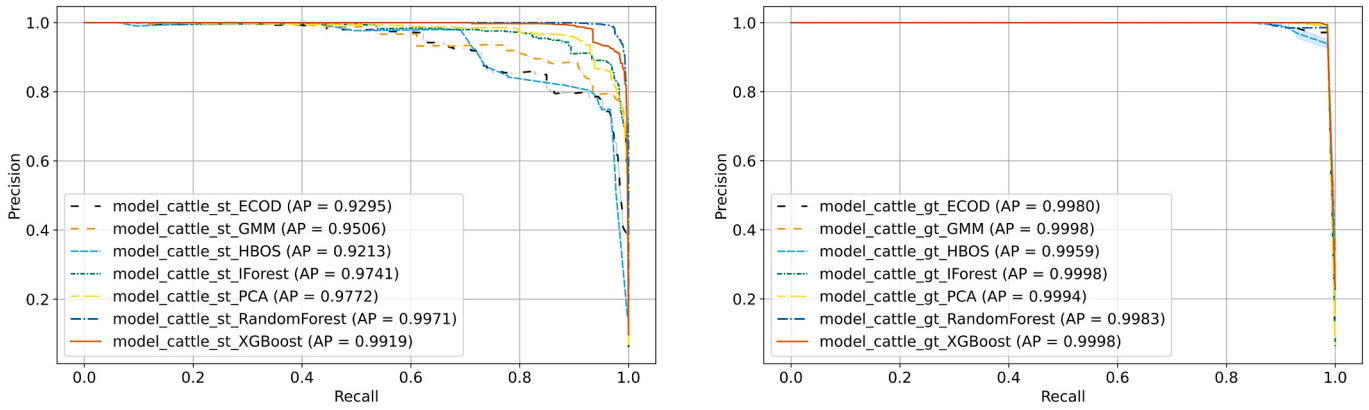


Fig. 3. Precision-recall curves of each model divided into single treatment (ST) on the left and group treatments (GT) on the right. AP is the average prescrion under the precision-recall cruve.

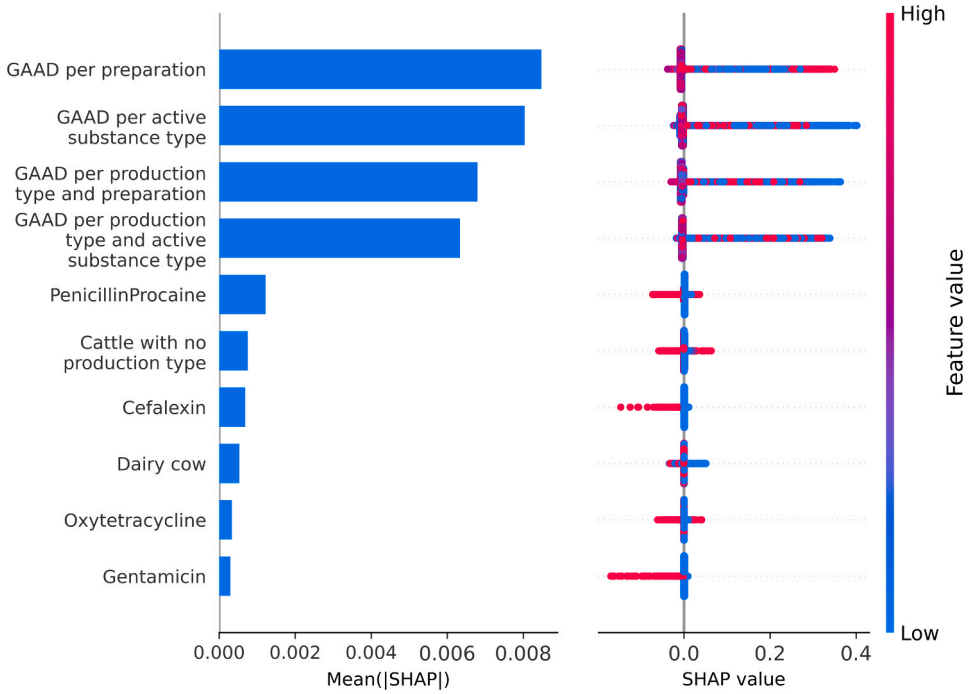


Fig. 4. The ten highest-contributing variables of ST Random Forest, determined through SHapley Additive Explanations (SHAP) values. On the left, a bar plot displays the mean absolute SHAP-values per variable. On the right, a SHAP beeswarm plot is shown where each dot represents an observation. Positive SHAP values indicate the change in the model's prediction towards an anomaly. Color indicates the original value of the variable; red indicates a high value and blue a low value. Variables are ranked according to the mean absolute SHAP values for all observations.

applying calibration with isotonic regression slightly improved ST Random Forests' brier and log loss to 0.000343 (-0.0003) and 0.001232 (-0.0019). Isotonic regression also increased the precision of ST Random Forest to 0.977 (+0.0391) and the F1-Score to 0.978 (+0.0149) while it decreased the ROC-AUC to 0.989 (Supplement Table 7). For the GT XGBoost we improved the brier and log loss to 0.00018 (-0.0013) and 0.00065 (-0.0336) respectively (Supplement Table 8). Isotonic regression increased the precision and the F1-Score to 0.986 (+0.0137) and to 0.993 (+0.0070).

4. Discussion

This study demonstrates that implementing the proposed two-staged anomaly detection framework can identify novel anomalies as well as reliably detect future outliers in data collected with the Swiss veterinary antibiotic surveillance system. Introducing population scaling transformations led to high F1-Scores on unseen data for single and group treatments. Both supervised models showed high rates of precision and recall across different thresholds and treatment types. The modified *given amount per animal and day* (GAAD) *per preparation* variable was among the top contributing variables for the identification of anomalies.

Similar to the Prescribed Daily Dose (PDD) (Merlo et al., 1996), the *given amount per animal and day* (GAAD) considers the actual amounts of prescribed medication rather than recommended doses. This variable enables us to detect potential outliers or deviations from the actually prescribed dosages seen in Swiss practices. Applying a logarithmic transformation to GAAD also helps align the scale by compressing larger values and the z-transformation standardizes the data by expressing each data point in terms of its distance from the mean in standard deviations (Kumpulainen et al., 2009). Prescriptions exhibiting extreme z-scores on any level are anticipated outliers that deviate considerably

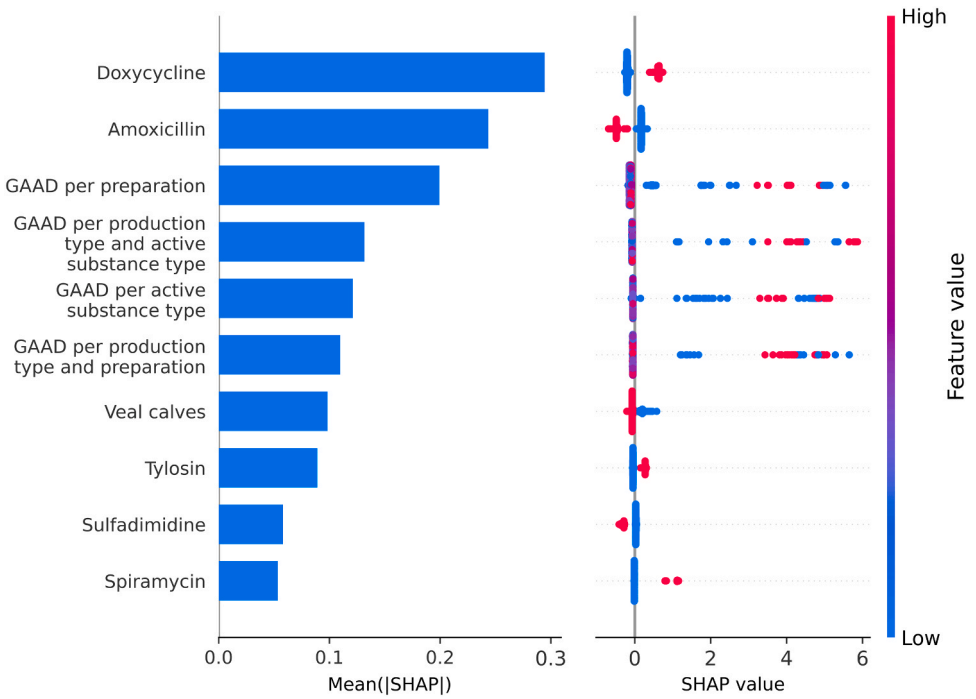


Fig. 5. The ten highest-contributing variables of GT XGBoost, determined through SHapley Additive Explanations (SHAP) values. On the left, a bar plot displays the mean absolute SHAP-values per variable. On the right, a SHAP beeswarm plot is shown where each dot represents an observation. Positive SHAP values indicate the change in the model's prediction towards an anomaly. Color indicates the original value of the variable; red indicates a high value and blue a low value. Variables are ranked according to the mean absolute SHAP values for all observations.

from the group mean.

The algorithms for the detection (Stage 1) were carefully selected to address the requirements of scaling to large datasets while maintaining speed. They were chosen to represent various types of detection methods, including linear model (PCA), probabilistic (GMM), probabilistic (ECOD), proximity-based HBOS, and ensemble-based Isolation Forest. Other common algorithms such as KNN, OCSVM, and LOF were initially considered but did not scale well with large amounts of data. OCSVM notoriously struggles with large datasets (Qiao et al., 2023), while KNN (Mucherino et al., 2009) and LOF (Breuniq et al., 2000), being proximity-based algorithms, were not able to effectively capture the distribution in the data. In the examples shown in Fig. 2 we observe that IForest ST and GMM GT draw different decision boundaries. However, we see in all plots (Supplement Fig. 1 and Supplement Fig. 2) a lowered outlier probability around the center (0,0) and increased outlier probability the further away the points are from the center. This scattering shows that all selected algorithms draw the desired decision boundary further away from the center. The combination of probabilities and votes ensures the continuity of the anomaly scores while allowing for a consensus-level evaluation. The results of the implemented voting classifier presented a list and a ranking of potential anomalies. This order enabled us, in collaboration with the experts, to identify previously undetected anomalies and define anomaly thresholds for each active substance, preparation, and production type manually.

The category "cattle with none of the specified production types" had the relative highest number of prescriptions with anomalies and was the production type with the 2nd highest number of prescriptions. Implemented as a residuary category, the frequent use of this group is indicative of its use instead of the specified production types. We suspect that veterinarians who do not care about entering the correct production type are also more likely to make also other inputs in a careless way. In contrast, in all specified production categories, the percentage of prescriptions with anomalies was almost level (0,82–1.31). The product category with the highest error rate was "tablets, capsules, and bolus" where veterinarians presumably prescribed a whole pack containing multiple e.g. tablets but just recorded a "One" as count, meaning one tablet. This might be a handling error or a transfer error of their practice software which unit they report (input variant 1). Although prescriptions of the category "premix for medicated feeding" are almost entirely entered directly in the IS ABV form (input variant 2 mandatory for oral group treatments) and thus no transfer errors can occur, the category has the 2nd highest percentage of anomalies. This is most likely because it contains the most input fields and thus, if a veterinarian makes a mistake in any of this fields, it can quickly falsify the GAAD of this prescription. An example of this error is when a veterinarian prescribes an entire premix for medicated feeding as a single treatment for one animal. Both types of udder injectors had the least relative number of anomalies given that data input, application and indication of this preparation type is most straightforward. For injection preparations veterinarians often did not state the appropriate number of animals treated, which led to a very high GAAD. For the active ingredient classes macrolides, especially Tulathromycin (2.3 %) and Tylosin (1.8 %) preparations had the most anomalies.

Reasons for anomalies could either be the amount of antibiotics, the duration of the treatment or the number of animals treated. With prescription data alone we can state that their ratio differs substantially from what is common, compared to other prescriptions and expert expectations. We and the experts at the FSVO suspect that a mistake made by many veterinarians and veterinary nurses is to keep the number of animals treated at one, albeit the actual number is higher. The default setting for the number of animals treated in single treatments is one and if unadjusted in the system could potentially lead to errors in the prescription. Another possibility is that the anomalies are simple data entry errors, which could be the result of inadequate staff training or simple negligence of the veterinarian. Anomalies may also represent systematic errors in data collection or transmission. However, in most cases, it is neither possible nor feasible to identify the main reason for each anomalous prescription from the data alone. Irrespective of the nature of the anomaly, if it is sufficiently different from the Swiss average, the

prescription should be flagged as an anomaly.

While the unsupervised method is great for discovery, the anomaly cutoff (vote) is based on the contamination factor, which is an inherently flawed concept for a planned live detection of anomalies in the IS ABV prescription system and can be difficult to determine (Perini et al., 2022). The number of incorrect prescriptions varies from year to year, and relying on a fixed percentage and an unsupervised approach it is likely to lead to under- or over-estimation of anomalies over time. To address this, our approach incorporates a supervised step to ensure more accurate and adaptive anomaly detection. Both supervised models (ST Random Forest and ST XGBoost) demonstrate a clear advantage over the unsupervised methods (Fig. 3). HBOS was the most sensitive to the different thresholds of the Precision-Recall curves. No clear difference is visible in the other algorithms for the Precision-Recall curves in group treatments. It is important to note, that input variant 2 must be used to send prescriptions of oral group therapies. Thus, over all input quality seems much better when every input is checked and directly entered in the IS ABV form. We find a strong difference in the separation between anomalies of single treatments and the anomalies of group treatments, which are much more distant and do not transition seamlessly (Fig. 2). This clear data separation benefits the GT algorithms in detecting anomalies and raises the question if such a machine learning approach is even necessary for group treatments in cattle or if a simple benchmark limit does the job.

In ST Random Forest we observe that the four most contributing variables are all the GAAD groups (Fig. 4). The GAAD per preparation had the most impact on ST Random Forest. Interestingly it ranked higher in ST Random Forest and in GT XGBoost than GAAD per production type and preparation. We would expect that the highest degree of information (GAAD per production type and preparation) would provide the most impact. GAAD per active substance type is a broader comparison of active substances and its high impact suggests that the population comparison with the different GAAD groups helps to detect anomalies. After the GAAD groups, the categorical variable Penicillin-Procaine is the 5th most important variable for ST Random Forest. Penicillin-Procaine is the active substance with the most anomalies (7747). Cattle with no production type and dairy cows are the most prevalent production type in the single treatment. Their relatively larger sizes seem to have high impact in SHAP variable contribution. In XGBoost the GAAD groups were not the most important variables. Doxycycline and Amoxicillin had the most prescriptions (96 and 58) with anomalies in group treatments and had the highest impact on determining anomalies. With the reduced number of prescriptions and anomalies in group treatments, the categorical active substance seems to have more impact on the determination of anomalies than the GAAD groups.

In our approach, we gave more weight to the minority class in training, which biased the predictions towards the minority class (Supplement Fig. 3). Even after isotonic calibration the ST Random Forest has a bend towards the bottom right corner which indicates an overestimation of risks in ST Random Forest (Van Calster et al., 2019). For GT XGBoost we do not achieve a stable calibration as the curve has a sigmoidal shape (Supplement Fig. 4). GT XGBoost is underconfident at high probabilities and overconfident when predicting low probabilities. A slight overestimation is the preferred state of the models because it prioritizes the detection of true anomalies, even if this leads to more false positives but emphasizes sensitivity over specificity.

Even though stage 1 delivers a more streamlined way of anomaly annotation, it still requires a considerable amount of human resources and time. We think that just relying on the unsupervised algorithms or the subsequent voting classifier is not sufficient to reliably identify anomalies in a prescription-based system. We reckon that our approach provides a helpful structure and supplementary information for expert labeling. Significant time savings will be achieved once the supervised algorithms can be applied to the system on a regular basis. Our method will be used for future prescriptions from 2024 and will allow the FSVO to give veterinary practices rapid feedback on the quality of their submissions, so that they can act accordingly.

5. Conclusion

Our approach is effective in detecting antibiotic prescriptions that deviate significantly from expected patterns in the Swiss veterinary population. By identifying outliers, we can pinpoint potential errors, anomalies, or cases of inappropriate antibiotic use, enabling veterinarians and practices to take measures to correct their inputs. The proposed two-stage anomaly detection is a valuable and effective tool to prevent future prescription errors in the Swiss veterinary antibiotic surveillance system ISABV.

Funding

This project was funded by the FVSO under 714001836.

CRediT authorship contribution statement

Guy-Alain Schnidrig: Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Anaïs Léger:** Writing – review & editing, Validation, Investigation. **Heinzpeter Schwermer:** Writing – review & editing, Validation. **Rebecca Furtado Jost:** Writing – review & editing, Validation, Investigation. **Dagmar Heim:** Writing – review & editing, Supervision, Funding acquisition. **Gertraud Schüpbach-Regula:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare there is no conflict of interest.

Acknowledgments

We like to thank all employees at the FSVO which were involved in this project.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, we used “DeepLWrite” to give suggestions to rearticulate some of the more convoluted sentences. After using this tool, we reviewed and edited the content as needed and take full responsibility for the content of the publication.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.prevetmed.2024.106291.

References

Aggarwal, C.C., 2017. Outlier Analysis. Springer, Cham https://doi.org/doi.org/10.1007/978-3-319-47578-3.

Amruthnath, N., Gupta, T., 2018. A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. 2018 5th Int. Conf. Ind. Eng. Appl. ICIEA. pp. 355–361. https://doi.org/10.1109/IEA.2018.8387124.

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A.F., Nielsen, H., 2000. Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16, 412–424. https://doi.org/10.1093/bioinformatics/16.5.412.

Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. 30, 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2.

Breuniq, M.M., Kriegel, H.P., Ng, R.T., Sander, J., 2000. OF: Identifying density-based local outliers. SIGMOD Rec. (ACM Spec. Interes. Gr. Manag. Data) 29, 93–104. https://doi.org/10.1145/335191.335388.

Bronzwaer, S.L.A.M., Cars, O., Udo Buchholz, S.M., Goettsch, W., Veldhuijzen Jacob, L., Kool, I.K., Sprenger, M.J.W., Degener, J.E., 2002. A European study on the relationship between antimicrobial use and antimicrobial resistance. Emerg. Infect. Dis. 8, 278–282. https://doi.org/10.3201/eid0803.010192.

Caneschi, A., Bardhi, A., Barbarossa, A., Zaghini, A., 2023. The use of antibiotics and antimicrobial resistance in veterinary medicine, a complex phenomenon: a narrative review. Antibiotics 12. https://doi.org/10.3390/antibiotics12030487.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 13-17-Augu, 785–794. https://doi.org/10.1145/2939672.2939785.

Daga, A.P., Fasana, A., Garibaldi, L., Marchesiello, S., 2020. On the use of PCA for diagnostics via novelty detection: interpretation, practical application notes and recommendation for use. PHM Soc. Eur. Conf. 5, 13. https://doi.org/10.36001/phme.2020.v5i1.1241.

European Medicines Agency, 2013. Revised ESVAC reflection paper on collecting data on consumption of antimicrobial agents per animal species, on technical units of measurement and indicators for reporting consumption of antimicrobial agents in animals. EMA/286416/2012-Rev.1 44, pp. 1–29..

FSVO, 2020a. ARCHVET. Report on the distribution of antibiotics and antibiotic resistance in veterinary medicine in Switzerland.

FSVO, 2020b. Jahresbericht IS ABV, Erste Übersicht der Verschreibungen von Antibiotika bei Nutztieren in der Schweiz.

Giedraitienė, A., Vitkauskienė, A., Naginienė, R., Pavilonis, A., 2011. Antibiotic resistance mechanisms of clinically important bacteria. Medicina 47, 19. https://doi.org/10.3390/medicina47030019.

Goldstein, M., Dengel, A., 2012. Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm. KI-2012 Poster Demo Track 59–63..

Goossens, H., Ferech, M., Vander Stichele, R., Elseviers, M., 2005. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. Lancet 365, 579–587. https://doi.org/10.1016/S0140-6736(05)17907-0.

Ho, T.K., 1995. Random decision forests Tin Kam Ho perceptron training. In: Proc. 3rd Int. Conf. Doc. Anal. Recognit. 1, 278–282..

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., 2016. Jupyter Notebooks—a publishing format for reproducible computational workflows. Position. Power Acad. Publ. Play. Agents Agendas - Proc. 20th Int. Conf. Electron. Publ. ELPUB 2016. pp. 87–90. https://doi.org/10.3233/978-1-61499-649-1-87.

Kumpulainen, P., Kylväjä, M., Hätönen, K., 2009. Importance of scaling in unsupervised distance-based anomaly detection.

Laxminarayan, R., Duse, A., Wattal, C., Zaidi, A.K.M., Wertheim, H.F.L., Sumpradit, N., Vlieghe, E., Hara, G.L., Gould, I.M., Goossens, H., Greko, C., So, A.D., Bigdeli, M., Tomson, G., Woodhouse, W., Ombaka, E., Peralta, A.Q., Qamar, F.N., Mir, F., Kariuki, S., Bhutta, Z.A., Coates, A., Bergstrom, R., Wright, G.D., Brown, E.D., Cars, O., 2013. Antibiotic resistance-the need for global solutions. Lancet Infect. Dis. 13, 1057–1098. https://doi.org/10.1016/S1473-3099(13)70318-9.

Li, Z., Zhao, Y., Hu, X., Botta, N., Ionescu, C., Chen, G., 2022. ECOD: Unsupervised outlier detection using empirical cumulative distribution functions. IEEE Trans. Knowl. Data Eng. 1–13. https://doi.org/10.1109/TKDE.2022.3159580.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 4766–4775.

Lynn, P.A., 1986. The laplace transform and the z-transform. in: Electronic Signals and Systems. Macmillan Education UK, London, pp. 225–272. https://doi.org/10.1007/978-1-349-18461-3_6.

Merlo, J., Wessling, A., Melander, A., 1996. Comparison of dose standard units for drug utilisation studies. Eur. J. Clin. Pharmacol. 50, 27–30. https://doi.org/10.1007/s002280050064.

Morales, F.A., Ramírez, J.M., Ramos, E.A., 2020. A mathematical assessment of the isolation tree method for data anomaly detection in big data. https://doi.org/https://doi.org/10.48550/arXiv.2004.04512.

Mucherino, A., Papajorgji, P.J., Pardalos, P.M., 2009. k-Nearest Neighbor Classification. Data Mining in Agriculture. Springer New York, New York, NY, pp. 83–106. https://doi.org/10.1007/978-0-387-88615-2_4.

Niculescu-Mizil, A., Caruana, R., 2005. Predicting good probabilities with supervised learning, In: Proceedings of the 22nd International Conference on Machine Learning, ICML '05. Association for Computing Machinery, New York, NY, USA, pp. 625–632. https://doi.org/10.1145/1102351.1102430.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Perini, L., Vercruyssen, V., Davis, J., 2022. Transferring the Contamination Factor between Anomaly Detection Domains by Shape Similarity. In: Proc. 36th AAAI Conf. Artif. Intell. AAAI 2022 36, 4128–4136. https://doi.org/10.1609/aaai.v36i4.20331.

Perry, J., Waglechner, N., Wright, G., 2016. The prehistory of antibiotic resistance. Cold Spring Harb. Perspect. Med. 6, 1–8. https://doi.org/10.1101/cshperspect.a025197.

Platt, J., 2000. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Adv. Large Margin Classif. 1, 10.

Prestinaci, F., Pezzotti, P., Pantosti, A., 2015. Antimicrobial resistance: a global multifaceted phenomenon. Pathog. Glob. Health 109, 309–318. https://doi.org/10.1179/2047773215Y.0000000030.

Qiao, Y., Wu, K., Jin, P., 2023. Efficient anomaly detection for high-dimensional sensing data with one-class support vector machine. IEEE Trans. Knowl. Data Eng. 35, 404–417. https://doi.org/10.1109/TKDE.2021.3077046.

van Rossum, G., 1995. Python tutorial, Technical Report CS-R9526. Cent. voor Wiskd. En. Inform.

RStudio Team, 2015. RStudio: integrated development environment for R.

Sanders, P., Vanderhaeghen, W., Fertner, M., Fuchs, K., Obritzhauser, W., Agunos, A., Carson, C., Borck Høg, B., Dalhoff Andersen, V., Chauvin, C., Hémonic, A., Käsbohrer, A., Merle, R., Alborali, G.L., Scali, F., Stärk, K.D.C., Muentener, C., van Geijlswijk, I., Broadfoot, F., Pokludová, L., Firth, C.L., Carmo, L.P., Manzanilla, E.G., Jensen, L., Sjölund, M., Pinto Ferreira, J., Brown, S., Heederik, D., Dewulf, J., 2020. Monitoring of farm-level antimicrobial use to guide stewardship: overview of existing systems and analysis of key components and processes. Front. Vet. Sci. https://doi.org/10.3389/fvets.2020.00540.

Sarker, I.H., 2021. Machine learning: algorithms, real-world applications and research directions. SN Comput. Sci. 2, 1–21. https://doi.org/10.1007/s42979-021-00592-x.

Shyu, M.L., Chen, S.C., Sarinnapakorn, K., Chang, L., 2003. A novel anomaly detection scheme based on principal component classifier. In: Proceedings of the 3rd IEEE Int. Conf. Data Min. 353–365..

Van Calster, B., McLernon, D.J., Van Smeden, M., Wynants, L., Steyerberg, E.W., Bossuyt, P., Collins, G.S., MacAskill, P., Moons, K.G.M., Vickers, A.J., 2019. Calibration: The Achilles heel of predictive analytics. BMC Med. 17, 1–7. https://doi.org/10.1186/s12916-019-1466-7.

Van De Sande-Bruinsma, N., Grundmann, H., Verloo, D., Tiemersma, E., Monen, J., Goossens, H., Ferech, M., 2008. Antimicrobial drug use and resistance in Europe. Emerg. Infect. Dis. 14, 1722–1730. https://doi.org/10.3201/eid1411.070467.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A., Pietro, Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D.V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. ⟨https://doi.org/10.1038/s41592-019-0686-2⟩.

Vovk, V., 2015. The fundamental nature of the log loss function. Lect. Notes Comput. Sci. (Incl. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.) 9300, 307–318. https://doi.org/10.1007/978-3-319-23534-9_20.

Winkler, R.L., 1994. Evaluating probabilities: asymmetric scoring rules. Manag. Sci. 40, 1395–1405.

Zhang, E., Zhang, Y., 2009. Average precision. In: LIU, L., ÖZSU, M.T. (Eds.), Encyclopedia of Database Systems. Springer US, Boston, MA, pp. 192–193. https://doi.org/10.1007/978-0-387-39940-9_482.

Zhao, Y., Nasrullah, Z., Li, Z., 2019. PyOD: A python toolbox for scalable outlier detection. J. Mach. Learn. Res 20, 1–7.