Estimation of Difficulty in Calving for Cows

Subin Sam Varghese
samvar01@gw.uni-passau.de
University of Passau

Rohit Ramakrishnan
ramakr02@gw.uni-passau.de
University of Passau

Siddhey Ratnaparkhi
ratnap01@gw.uni-passau.de
University of Passau

Saikrishna Dhiddi
dhiddi01@gw.uni-passau.de
University of Passau

ABSTRACT

For better yield and productivity of diary farms, it is important to monitor and predict the nature of Calving process in cows. This helps in taking preventive measures during a possible difficult calving, resulting in healthier calves, cows and prevent loss of cattle lives.

Naturally, in order to prepare for a complicated calving scenario, the nature of calving has to be predicted well before the actual instance of birth. For prediction of the nature of calving, we use sensor data for monitoring various activities of the cow and other information such as the race of cow, number of prior births and Temperature-Humidity Index etc. along with the recorded difficulty level of the calving that cow had. This problem is an example of supervised learning, where we model a relation between the activity data from sensors and other general data of a cow to the difficulty level of calving for a cow. The inference of the statistical analysis and machine learning can help in predicting the nature of calving and can help in setting up preventive medical assistance in case of a complicated calving. The report includes how the difficulty levels are mapped to data, machine learning approaches used and other factors which has to be considered for the practical usage of a model, such as cost associated for setting up external assistance and probability of incorrect prediction of difficult calving.

INTRODUCTION

Sensor data is obtained by placing sensors in the location (or) on the object, which detects and collects the data from the changes(of the environment) or action(of objects). It can be further analyzed and then respective measures can be taken. In this experiment, we analyze sensor data from cows, with sensors attached at different points of time when a cow leaves the herd. The original dataset is provided by the **LFL Bayern (Bayerische Landesanstalt fuer Landwirtschaft)**.

The dataset consists of sensor values collected from more than sixty cows from July 2018 to January 2019 using a stomach sensor which measures the activity and stomach temperature every 10 minutes and a collar sensor which collects the rumination data of cow for every 2 hours.

Profound knowledge on the behavioural changes of the animal can be gained from the data by analyzing with different perspectives and more attention around calving period. The data trends around the calving period helps us to predict the difficulty of calving and also to identify time periods with a special need for observation on cow[1].

In the research work carried by Cameron E. F. Clark et al. 2014 [2] heat and rumination sensors were used for predicting the calving events. The research concluded that decrease in the duration of rumination can be used to predict the instance of calving. It also stated that further research is required whether rumination data and level of activity can predict if external assistance in calving is required or not.

The main objectives of this experimental study are:

- To analyze the activity, rumination and temperature variations before calving.
- To use statistical analysis and machine learning techniques on the provided data.
- To predict the difficulty of calving.

1 PROBLEM STATEMENT

Based on the literature available on calving prediction and behaviour of cows during the period of calving, it has been illustrated that the combination of cow activity, rumination time and lying behavior data can be used to provide insights on the calving prediction. The rumination time around the period of calving has been observed to be reduced.[1].

The aim of this project is to estimate the difficulty in calving based on the analysis of rumination, activity and temperature data. The inference can be evolved using statistical and machine learning techniques on the given data. Using features which corresponds to an activity, a classifier can be trained for the estimation of distinct difficulty levels in calving. The distinct levels of difficulty, provided in the dataset, are used as reference classes for the extent of difficulty in calving. The levels are numbered from 0 to 4. 0 is the least difficult calving and 4 is the most difficult calving. The mapping of extracted features and corresponding level of difficult calving would be used for training the classifier.

The generated classifier can be then used to predict the level of difficult calving from the activity attributes which were used to train the classifier. Later, the K-fold validation method would be used to evaluate the trained classifier and measure the performance of the classifier.

1.1 Approach

The phases of the project can be categorized into three parts as follows:

- Analysis of Data
- Model Implementation

Data Science Lab, 2019, Uni Passau
.

- Evaluation

The flowchart in Figure 1 below describes the workflow of the project.

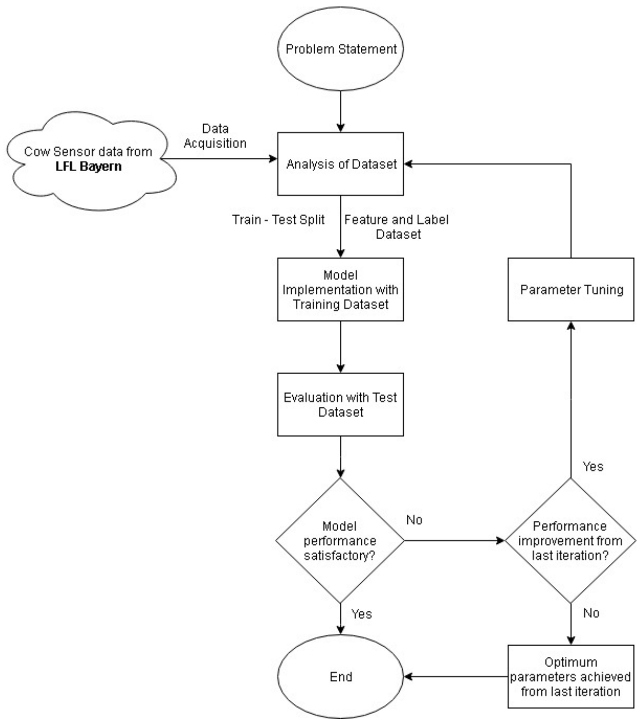

Figure 1: Project Flowchart

1.1.1 Analysis of Data.
The dataset consists of rumination and activity data, collected from collar sensor, measured for two hours where as a stomach sensor provides temperature and stomach-activity data measured for every ten minutes, daily for each cow. Alongwith cow activty data, we also have data regarding the environment such as the Temperature-Humidity Index. Information on the race of the cow, number of prior births and if the calves were twins or not is provided.

The recorded data consists of the attributes collected before and after the calving process. However, the problem statement focuses on the difficulty in calving. Thus, the train data is formed considering only the attributes before and at the time of calving. By looking into the trends of rumination data, cow activity and temperature data corresponding to the difficulty of calving could help in forming a prediction model.

The calving difficulty level data scaled from 0 to 4 can be grouped into 3 classes. The difficulty level 0 of no external help calving and difficulty level 1 of easy calving can be combined into one class of easy calving. The difficulty level 2 can remain as a single class of moderate calving whereas the remaining levels of 3 and 4 can be combined into a third class of difficult calving.

1.1.2 Model Implementation.
Machine Learning techniques can be used to predict the difficulty in the calving process. According to the dataset the level of difficulty is divided into 5 levels(0-4).The proposed classifier classifies the difficulty in calving into 3 classes. Class 1 of easy calving, Class 2 of moderate calving and Class 3 of difficult calving. Some of the algorithms which can be used for predicting the nature of calving process are Neural Network, Linear Discriminant, Random Forest and K-Nearest Neighbours(KNN) [1].

1.1.3 Evaluation.
The evaluation of the classifier is conducted based on K-fold cross validation. The suitable value of K would be found based on the size of the processed input dataset. The performance metrics would be evaluated based on the ordinal classes of calving difficulty, since a relative ordering between Easy, Medium and Difficult calving exists.

2 DATA ACQUISITION & PRE-PROCESSING

The data collected for the entire set of cows is varied and provides several attributes, which could offer potential features to be processed and extracted, later fed as an input to the classifier model. This phase of the report focuses on providing information over the acquisition of data, data preprocessing and feature engineering. The following subsections elaborate over the same in detail.

2.1 Data acquisition

The governing body 'Bayerisches Staatsministerium fuer Ernaehrung, Landwirtschaft und Forsten' is the ministry of free state of Bavaria, performing research at institute LFL Bayern (Bayerische Landesanstalt fuer Landwirtschaft) which offers the required sensor data collected over a fixed set of cows. The sensors used to acquire the data are provided by 'smaXtec' which deals with monitoring the health status of the animal and calving prediction. The 'smaXtec Classic Bolus' sensor provides measurement data for the attributes of stomach activity and stomach temperature. The sensor data is provided in the form of Spreadsheets which includes distinct columns of attributes. This acquired data will now be used for data preprocessing.

To visualize the three distinct levels of calving difficulty, the *Principle Component Analyzer* from the Scikit-Learn has been used.[1]

2.2 Data preprocessing

2.2.1 Filtering, Grouping and Statistics.
The required data columns are filtered over the grouped difficulty levels of calving scaled as 0, 1, 2, 3 and 4 as shown in Table 1. As the difficulty level 3 does not have sensor data attributes of rumination and activity whereas there exists only one cow with difficulty level 4, these levels are considered as outliers and do not count in the further preprocessing tasks. Thus, the groups of difficulty levels are 0, 1 and 2.

The filtering process also involves the cleaning of data. The data set involves many Missing Values and data entries which has 'NaN' (Not a Number) as a data entry in the cells. We drop these entries to clean the dataset. This improves the correlation score amongst the considered attribute and the difficulty level. The aim of correlation

[1] https://scikit-learn.org/stable/

	Kuhnummer	Reine Aktivitätsdaten	Rohdaten Rumination
count	1.038898e+06	179617.000000	179587.000000
mean	1.050592e+12	44.877873	44.470207
std	1.650958e+13	15.128753	21.514714
min	0.000000e+00	5.000000	0.000000
25%	1.560000e+02	34.000000	30.000000
50%	4.119500e+04	42.000000	45.000000
75%	4.238900e+04	53.000000	60.000000
max	2.760009e+14	168.000000	138.000000

Figure 2: Statistical data on attributes including Rumination and Cow activity

	Nummer	Laktation	Kalbeverlauf	Unnamed: 16	Lakt.-Tage
count	126.000000	126.000000	120.000000	0.0	86.000000
mean	387.817460	2.047619	0.791667	NaN	284.604651
std	377.547653	2.058571	0.994910	NaN	74.768235
min	2.000000	0.000000	0.000000	NaN	2.000000
25%	85.250000	0.000000	0.000000	NaN	222.750000
50%	158.500000	2.000000	0.000000	NaN	288.000000
75%	838.000000	3.000000	1.000000	NaN	344.000000
max	999.000000	8.000000	4.000000	NaN	439.000000

Figure 3: Statistical data on attributes including Lactation days and Calving difficulty levels

Table 1: Total Number of Cows in each Difficulty Level

Difficulty Level	Number of Cows
0	61
1	35
2	13
3	10
4	1

is to find a significant score for the relevant data with the Difficulty Level.

For instance the relevant columns are:

(1) Lactation
(2) Lactation Days
(3) Stomach Activity
(4) Stomach Temperature

Correlation scores of these relevant columns with Difficulty Level:

(1) 0.182484
(2) -0.225164
(3) 0.179419
(4) 0.005060

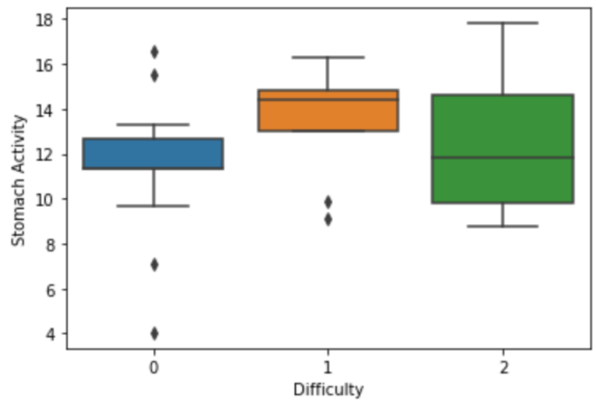

Figure 4: Stomach Activity plotted over different calving difficulty levels

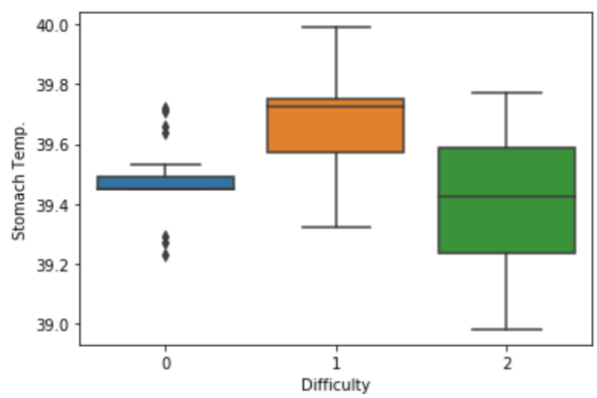

Figure 5: Stomach Temperature plotted over different calving difficulty levels

The same would be followed for other relevant columns such as Rumination and Pure Activity data. In general, the positive and negative correlation values help to map its behaviour to the corresponding classes of difficulty levels 0, 1 and 2. From Figure 2 and Figure 3, the High and Low Activity groups can be recorded which consists of cows with all the difficulty levels. The High Rumination value recorded is of 138 and lowest is 0. Also, the Highest Laktation recorded is of value 8 and lowest is 0. The overall trend observed says that the Cows with groups of low activity of rumination at the time of calving has a certainty of having moderate or difficult calving. The high Laktation group shows high certainty for a moderate or difficult calving.

As the dataset after cleaning and treating the outliers is still vast and is hard to visualize in terms of finding a suitable trend or information, a small subset of the given dataset is considered for the next step of feature engineering. Figure 15 and Figure 16 shows thirty six datapoints with corresponding features used to plot the

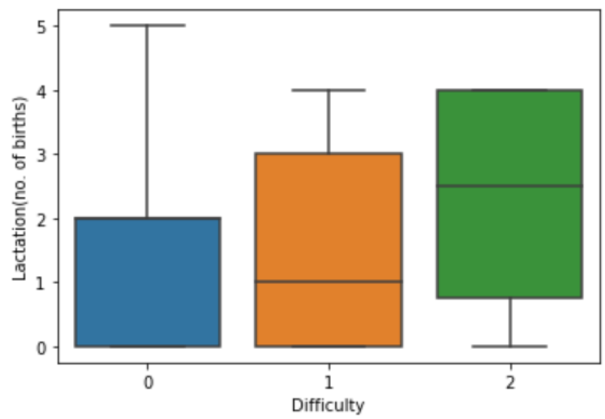

Figure 6: Lactation (number of prior Births) plotted over different calving difficulty levels

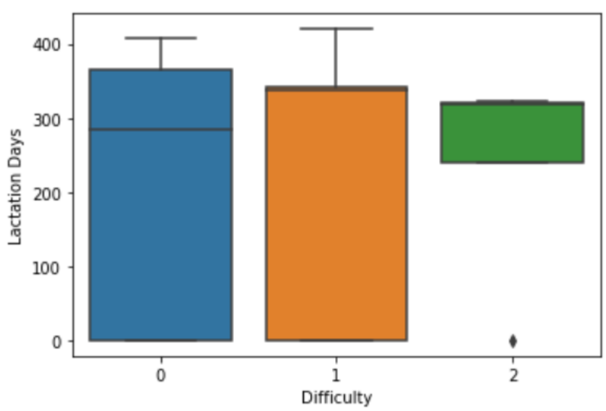

Figure 7: Lactation Days plotted over different calving difficulty levels

Table 2: Total Number of Cows in each Difficulty Level after Preprocessing

Difficulty Level	Number of Cows
0	19
1	9
2	4

PCA.

2.3 Feature engineering

2.3.1 Feature transformation. The extracted features are the filtered data subsets (unique number of cows) obtained from the earlier step of data preprocessing. The following data in Table 2 provides the information on the data subset of unique cows.

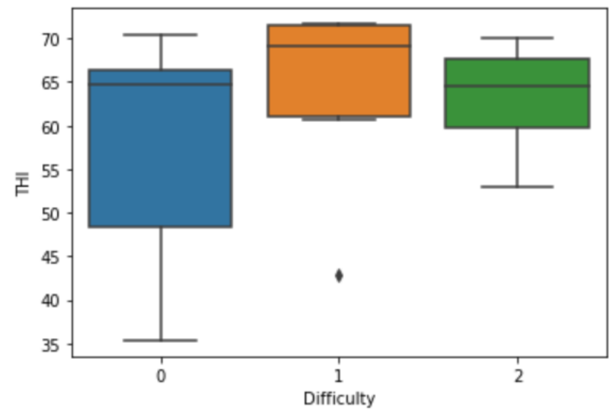

Figure 8: THI (Temperature Humidity Index) plotted over different calving difficulty levels

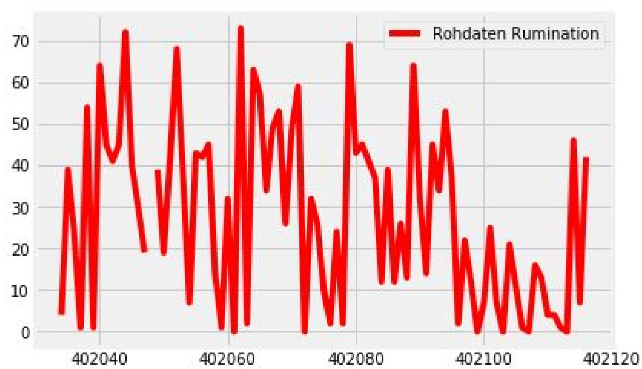

Figure 9: Rumination data plotted for calving difficulty level 0

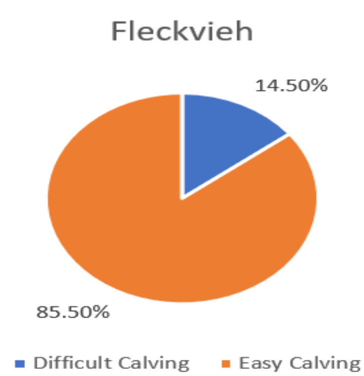

Figure 10: Probability of Easy or Difficult Calving based on Race-Fleckvieh of the Cow

Feature transformation is necessary as the distinct relevant feature subsets are categorized over three difficulty levels of calving

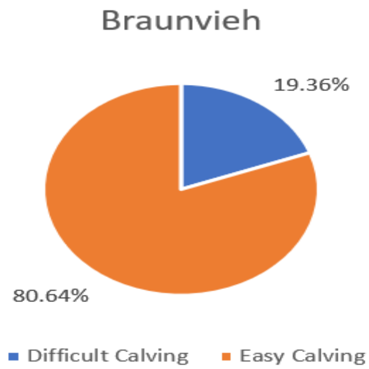

Figure 11: Probability of Easy or Difficult Calving based on Race-Braunvieh of the Cow

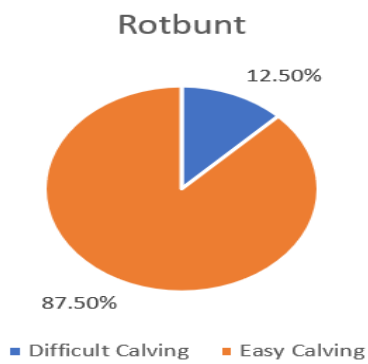

Figure 12: Probability of Easy or Difficult Calving based on Race-Rotbunt of the Cow

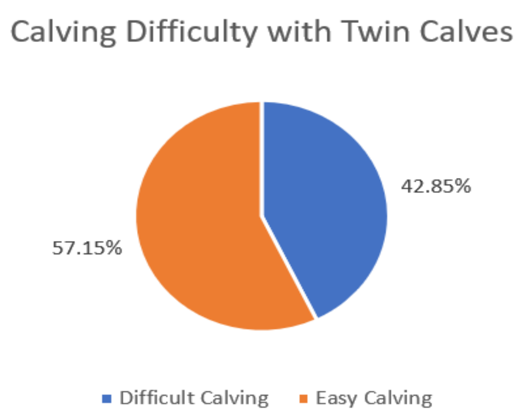

Figure 13: Probability of Easy or Difficult Calving based on Cow having Twin Calves

viz. Level 0, Level 1 and Level 2. In order to have a clear visualization of the feature subsets as categorized data, the methods of data analysis provides a visual aid. The methods considered to offer better visualization are Principal Component Analysis (PCA) and

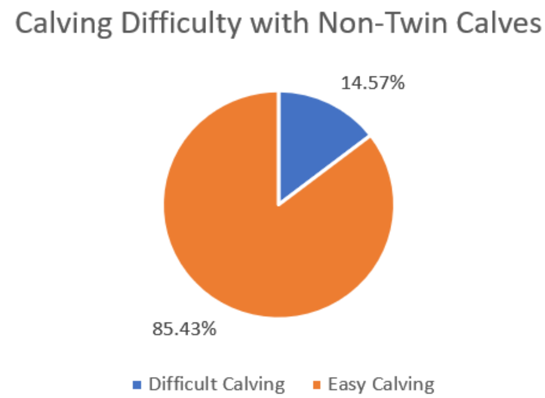

Figure 14: Probability of Easy or Difficult Calving based on Cow having Non-Twin Calves

Linear Discriminant Analysis (LDA). As the PCA method involves maximizing the variance of the data elements and distinct plotting of the feature subsets corresponding to the target class of difficulty levels, it would be a better choice over LDA. The PCA plot shown in Figure 17 illustrates color coded data points for the three distinct calving difficulty levels (Level 0: Blue, Level 1: Magenta, Level 2: Yellow). The approach of estimating the difficult calving with levels of 0 and 1 as one class and rest as Outliers and then classifying the level 0 and level 1 as a binary classifier seems to be a good option. However, as seen from the PCA plot, the datapoints are too close and less to implement this approach in an efficient manner. Thus, a better approach for the given dataset is to classify the three levels of difficult calving distinctly.

2.3.2 Feature Statistics. Another set of information is obtained from the features such as THI (Temperature-Humidity Index), Race of the Cow and Twin Calves. The following features are plotted with respect to its calving difficulty levels. Figure 4 shows the Boxplot for feature Stomach Activity where the median of stomach activity for Class 2 difficulty level is 12. Figure 5 shows that the cows with difficulty level 2 has stomach temperature values ranging from values 39.2 to 39.6. The ramp increase in the Lactation ('Laktation') as shown in Figure 6 shows that the cows which had higher number of lactation count (prior births) delivers higher levels of calving difficulty. The THI range for difficulty level 2 of the cows, as shown in Figure 8 varies on an average (data averaged over a period of one month before calving) in an interval of 60 to 67. The cyclic trend observed for rumination data as shown in Figure 9 is periodic and has lower values for data recorded for the nights.

The Race and Twin Calves features of the Cow plays vital role in providing a certainty of easy or difficult calving irrespective of the other features. Although the certainty could be assured by training the classifier model with other features being considered. The following pie charts are the illustrations of these features with respect to calving difficulty (easy: level 0 and level 1; difficult: level 2). As shown in Figure 11, the race 'Braunvieh' provides a probability of

	Difficulty	Cow Number	Stomach Activity	Stomach Temp.
0	0	4	13.161802	39.273685
1	0	986	12.999240	39.659030
2	0	58	12.298090	39.446530
3	0	915	7.128638	39.530470
4	0	135	4.042967	39.261500
5	0	930	9.663200	39.288970
6	0	999	16.570790	39.712460
7	0	784	9.849773	39.228610
8	0	138	13.312670	39.718240
9	0	141	15.516510	39.641330
10	1	846	9.089993	39.471770
11	1	69	9.868639	39.747550
12	1	937	14.508430	39.774970
13	1	5	15.285000	39.745350
14	1	139	14.050690	39.316820
15	1	144	12.999880	39.566580
16	1	924	16.272580	39.689480
17	1	143	14.820230	39.985140
18	2	72	13.543890	39.771950
19	2	888	17.801260	39.528900
20	2	874	8.773832	39.316810
21	2	155	10.158320	38.983110

Figure 15: PCA data with Cow activity and temperature for different Calving Difficulty levels

19.36% difficult calving. On a contrary, the race 'Rotbunt' provides the highest probability of 87.50% easy calving. Figure 13 with feature Twin Calves shows a high certainty of 42.85% difficult calving. In Figure 14, the non-twin calves feature shows a low probability of 14.57% difficult calving. The features showing higher probability of difficult calving irrespective of the other features are crucial in order to have preventive measures prior to resolving a scenario of difficult calving.

The complete features mentioned above would together form an input dataset for the classifier model to be trained in the next step of model implementation.

	Difficulty	Cow Number	Rumination avg	Lactation days
0	0	999	29.5000	359
1	0	162	145.0000	0
2	0	930	21.0000	340
3	0	163	45.0000	0
4	0	10	33.0000	284
5	0	22	48.0188	222
6	0	54	31.0000	408
7	0	135	29.5000	0
8	0	16	27.5000	258
9	0	161	34.5000	0
10	0	89	26.5000	208
11	0	845	38.5000	401
12	1	937	32.1428	338
13	1	5	47.7970	341
14	1	166	58.0000	0
15	2	72	44.5531	319

Figure 16: PCA data with Rumination activity and Lactation days for different Calving Difficulty levels

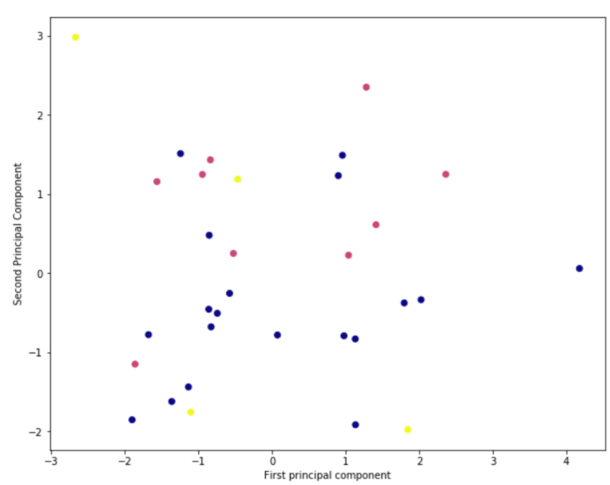

Figure 17: PCA scatter plot for 3 different Calving Difficulty levels

3 MODEL IMPLEMENTATION

3.1 Methodology

The goal of this section is to predict the difficulty of calving. The prediction is carried out by considering features obtained from second phase, and in addition to that, lactation(no. of births before observed calving) and Race as features are considered in this phase as shown in Figure 10. During feature engineering, correlation values observed for features such as 'Stomach activity', 'Race' and 'Number of prior births', were higher values than rest of the features. So, we have shortlisted six features as given below:

(1) Stomach Activity
(2) Lactation(Number of Prior births)
(3) Race
(4) Temperature of Cow
(5) Rumination
(6) Lactation days
(7) Twin Calves
(8) THI

	Difficulty	Cow Number	Stomach Activity	Stomach Temp.	Lactation(no. of births)	Rumination avg	Lactation Days	Rasse	Braunvieh	Fleckvieh	Rotbunt	Twin Calves	THI
27	0	845	11.320625	39.451500	4	38.5000	401	Fleckvieh	0	1	0	0	65.396334
28	1	937	14.510000	39.770000	3	32.1428	338	Fleckvieh	0	1	0	0	71.658819
29	1	5	15.290000	39.750000	2	47.7970	341	Fleckvieh	0	1	0	0	69.065863
30	1	166	14.387500	39.727917	0	58.0000	0	Braunvieh	1	0	0	0	42.942168
31	2	72	13.540000	39.770000	1	44.5531	319	Fleckvieh	0	1	0	0	70.085377

Figure 18: Initial data elements with all the selected features and difficulty level

The number of cows for different difficulty levels are unequal, hence we have an imbalanced dataset. In addition to that, the number of cows having all the features for this approach are less in number, we have 32 data points, due to this the computational time is considerably less. So, we have tried using multiple classifiers to observe their performance. Since the data is labeled and categorical, the classifiers used are SVM, Logistic Regression, Decision Tree, Naive Bayes and Random Forest. Once the model is trained, unseen test data is provided to the classifier to check the accuracy of prediction of calving difficulty. Table 3 represents the average accuracy values for different classifiers.

3.2 Experimental Setup

The Race feature consists of categorical data in the form of string. In order to use this feature as a part of training dataset, we use one hot encoding to handle the feature. Since the dataset is imbalanced with respect to the classes of difficulty levels, we choose stratified K-fold splitting to ensure all classes are represented in the train and test phases.

Our dataset contains imbalanced data and the choice of k value for implementing stratified K-fold has to be selected based on the results of the evaluation. The choice of K is critical and produces varying evaluation results. The evaluation metric (micro-precision, micro recall) yields better results when value of K is chosen as 8 over value of k as 15. We have recorded the accuracy for different number of folds[2] and for different classifiers in Table 2. From Table 2, it is observed that, as the value of K increases, the accuracy improves. However, for value of k equal to 8, the evaluation metrics of micro-recall and micro-precision results were observed to be better over value of k equal to 15.

[2]https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html

Table 3: Average Accuracy scores for different folds in K-fold validation rounded upto 4 decimal points

Classifier	K = 2	K = 5	K = 8	K = 10	K = 12	K = 15
Logistic Regression	0.4352	0.5713	0.5671	0.5870	0.6173	0.6473
SVM	0.6568	0.6572	0.6614	0.6642	0.6757	0.6928
Random Forest	0.6235	0.6235	0.6991	0.7210	0.7264	0.7396
Decision Tree	0.5647	0.7422	0.7654	0.7594	0.7977	0.8144
Naive Bayes	0.5607	0.6652	0.7108	0.7132	0.7252	0.7515

For choosing classifier we have referred to an online resource by Scikit Learn[3]. Based on the assumption that some features (lactation(Number of prior births), Stomach activity) play independent role in predicting the calving difficulty, Naive Bayes could be a suitable classifier for our task. Logistic regression works on the probabilities representing the possible outcomes of a single iteration which are modeled by using a logistic function. Also, one of the models used in Classification is Decision Tree. Considering our dataset, the complexity of the Decision Tree model will be less.

In the feature engineering phase, we used mean values to predict the missing data. Random Forrest consist of methods which helps to deal with the missing data without affecting the accuracy. In this phase, we used Random forest with the mean values. During the subsequent phase, we evaluate the prediction of missing data using Random Forest. In addition to the earlier mentioned classifiers, we also have observed the performance of SVM using our data, which is normally used on datasets with large number of features.

The results obtained for the Model Implementation will now be used by the means of Experiments in the last phase of Evaluation which will comment on the overall quality of the problem solution.

4 EVALUATION

4.1 Results

Since the dataset is imbalanced in each class as shown in Table 2, we have to look into the macro metrics such as macro-recall and macro-precision. This is significant to analyze the overall performance of the classifier for each class. Furthermore, we focus on achieving a higher recall value, as in the context of predicting a difficult calving, it is more important that a diary farmer is better prepared for a possible difficult calving which could be false than missing the detection of a difficult calving. Figure 19 to Figure 24 shows the values of Recall and Precision for three calving difficulty levels for different classifiers including a baseline classifier. The baseline classifier uses probability of each sample to predict the classes randomly.

All selected classifiers except Decision Tree (for class 0), has better recall values as compared to the precision values. As shown in figure 19, the Baseline model has a recall and precision score of 68% and 50.83% for class 0(Easy Claving Difficulty) respectively, 31%

[3]https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

Table 4: Micro Recall and Macro Recall for some classifier when the value of k=8 in K fold validation

Classifier	Macro Recall	Micro Recall
Baseline	40.28	56.25
Logistic Regression	44.10	52.10
SVM	52.08	66.71
Random Forest	75	75
Decision Tree	72.92	80.21
Naive Bayes	69.79	76.04

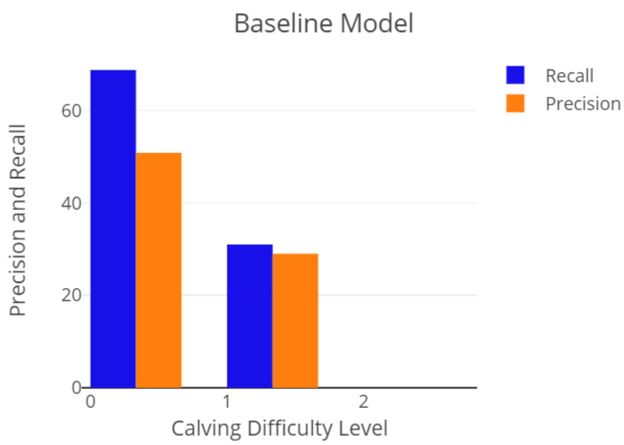

Figure 19: Recall and Precision for Baseline classifier

and 29% for class 1(Medium Calving Difficulty) and nil Precision and Recall for class 2. Since the number of data points in class 2 (Difficult Calving) are low(4 in quantity) compared to other two classes, the Baseline model is not able to classify this class. All other models evaluated in Phase 3 perform better than the Baseline model in every class.

Among the models used, Decision Tree performs the best. As observed in Figure 23, Decision Tree has a Recall and Precision score of 83.33% and 84.375% for class 0 respectively, 68.75% and 62.5% for class 1 and 50% and 31.2% for class 2. As shown in figure 25, Decision Tree assigns rules to classify the data, this results in a better classification of the dataset, which is sparse in our case. Random forest performs better in Recall score for class 0 and class 1 but does not detect class 2. Figure 24 shows that Random Forest has a Recall and Precision score of 93.75% and 79.16% for class 0 respectively, 81.25% and 68.75% for class 1 and nil Recall and Precision for class 2. Apart from Decision Tree, Naive Bayes performs better for class 2. Figure 22 shows the Recall and Precision values for Naive Bayes as 37.5% each for class 2. Figure 20 and figure 21 shows that the performance of the models Logistic Regression and Support Vector Machine are similar to the Baseline model.

4.2 Discussion

We have attempted to work on most of the features which is provided in the data set as shown in figure 18. In our case, the dataset

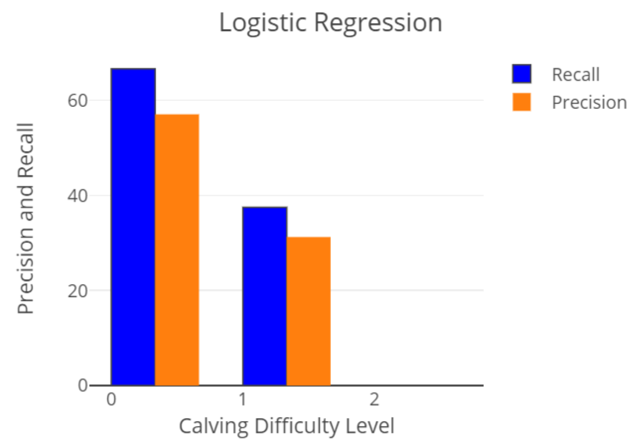

Figure 20: Recall and Precision for Logistic Regression classifier

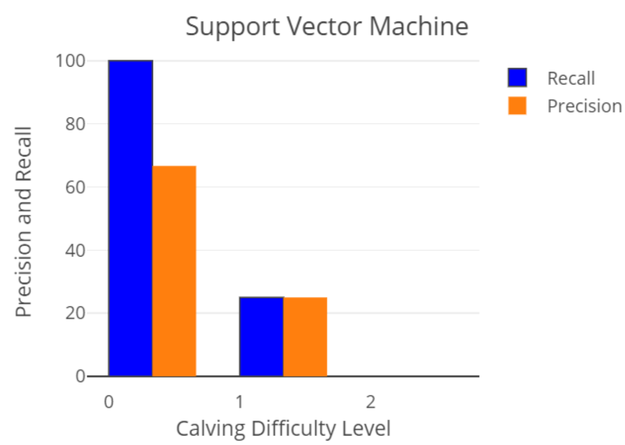

Figure 21: Recall and Precision for Support Vector Machine classifier

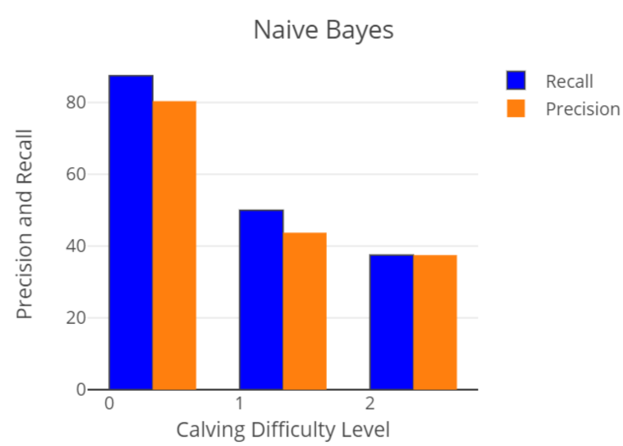

Figure 22: Recall and Precision for Naive Bayes classifier

Figure 23: Recall and Precision for Decision Tree classifier

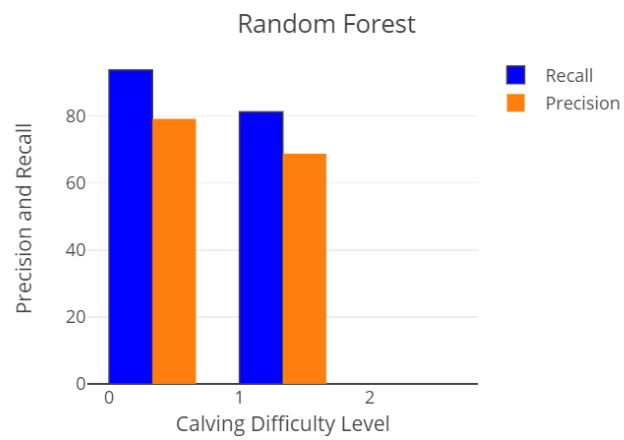

Figure 24: Recall and Precision for Random Forest classifier

provided does not have more than 20 objects in each class, nevertheless, few insights can be found which helps in the estimation of the difficulty in calving. Figure 25 is the visualization of the decision tree created using Sklearn library[4]. The tree depicts that class 0 can be predicted if the rumination average is less than 41.489 and THI value is less than 71.046 and also if the rumination average is greater than 44.77 and stomach activity is less than 12.854. Also, Class 2 can be predicted if rumination average is between 44.515 and 44.77 and rest of the cases are classified as class 1.

Analyzing overall data, Figure 7 to 9 represents the probability of Easy or Difficult calving based on the Race of cow. There are three races namely 'Braunvieh', 'Fleckvieh' and 'Rotbunt'. The data shows that for Race 'Braunvieh', the probability of a difficult calving is 19.36% which is more than that of 'Fleckvieh' and 'Rotbunt'. This analysis is helpful from the Farmers' perspective to be prepared for a difficult calving beforehand if the Cow is of Race

[4]https://scikit-learn.org/stable/modules/generated/sklearn.tree.export_graphviz.html

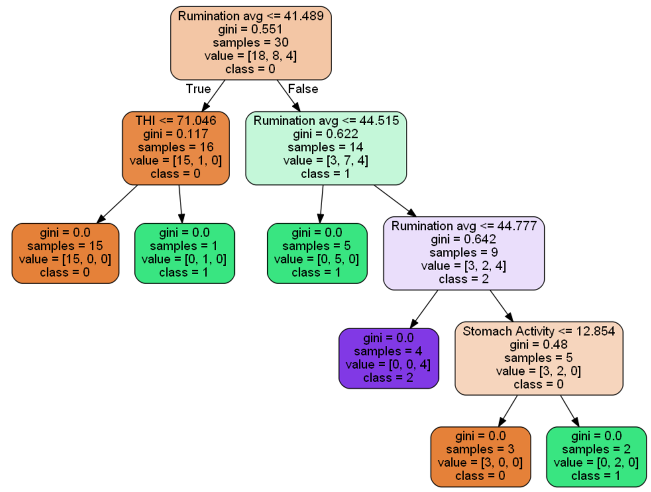

Figure 25: Visualization of Decision Tree

'Braunvieh' as the probability of having a difficult calving in future is more as compared to the other two races. In contrast to this, if the race of the cow is 'Rotbunt', the farmer would be in a place of assurance where the probability of having an easy calving is high.

In addition to this, Figure 10 to 11 shows the probability of Easy or Difficult calving based on cows having twin calves. From the data, it is observed that, in case of twin calves, the probability of difficult calving is 42.85% which is more than that of a non-twin calving. The non twin calving has a probability of 14.57% for difficult calving. Based on the analysis of the dataset the farmer can have prior knowledge on the percentage of cows which could have difficult calving with twin calves, hence farmer can be prepared to take preventive measures for approximate percentage of cows.

The estimation of difficulty in calving can thus be achieved distinctly with the help of other features using our trained models, amongst which Decision Tree is the most suitable model.

5 PHASE RESPONSIBILITY

(1) Rohit Ramakrishnan
(2) Siddhey Ratnaparkhi
(3) Saikrishna Dhiddi
(4) Subin Sam Varghese

REFERENCES

[1] Matthew. R. Borchers, Yu Mei Ruby Chang, Kathryn L. Proudfoot, B. A. Wadsworth, Amanda E. Stone, and Jeffrey M. Bewle. 2017. Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. *Journal of Dairy Science* 100, 7 (July 2017), 5664–5674. https://doi.org/10.3168/jds.2016-11526
[2] Cameron E. F. Clark, NA Lyons, L Millapan, S Talukder, GM Cronin, KL Kerrisk, and SC Garcia. 2015. Rumination and activity levels as predictors of calving for dairy cows. *animal* 9, 4 (2015), 691–695.