CP 218: Automotive Engine Analysis using Bayesian Learning

Project 2 Report:Team Data Miners

Md Farhad Hussain (23113) | Ashhar Zaman (22881) | Prashant Anil Jadhao (22432)

Abstract

A predictive maintenance model is build for automotive engines. Manual ways are present for people to check the engine health such as oil level and condition, coolant, brake fluid level etc. But to automate the process we can use the data of all the parameters available and make a system to predict the health of an automotive engine. By analyzing the patterns and trends in the data, bayesian machine learning algorithms are trained to predict when an engine is likely to require maintenance or repair. This could help vehicle owners and mechanics proactively address potential issues before they become more severe, leading to better vehicle performance and longer engine lifetimes. Other analysis such as linear regression and clustering is done using bayesian techniques.

1 Introduction

In the world of car maintenance, using smart computer techniques can help predict when engines need fixing before they actually break down. One way to do this is by using Bayesian learning approaches. These methods help us make guesses about what might happen in the future based on what's happened in the past.

Linear regression helps us predict engine problems by looking at data like engine temperature and RPMs. It's like connecting the dots to see where the engine might be heading. Classification helps us figure out if an engine is healthy or not. It's like sorting engines into "good" and "bad" groups.

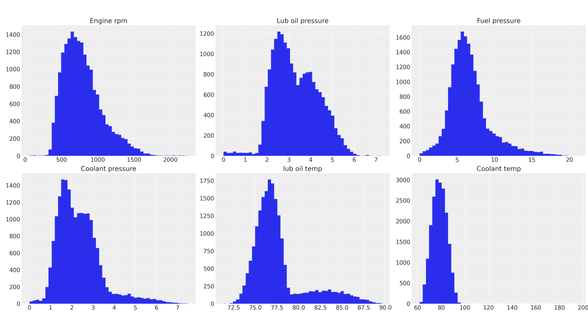
2 Dataset

The dataset includes various features and measurements related to the engine health of vehicles, such as engine RPM, Lub (lubricant) oil pressure, fuel pressure, coolant pressure, lub oil temp, coolant temperature, engine health.[1]

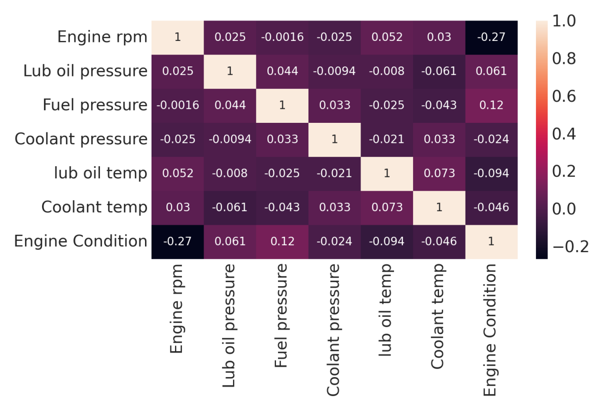
3 Data Pre-processing

3.1 Distribution Statistics of Features

To gain insights into the dataset, distribution statistics of features were computed. Visual aids, in the form of plots, were generated to facilitate both visualization and processing.



(a) Data distribution



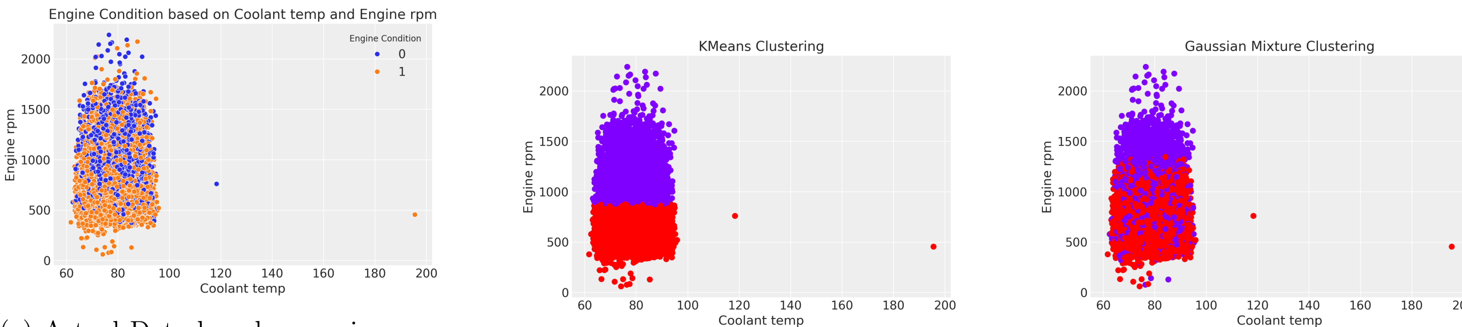
(b) Data Correlation

Figure 1: Data Visualization

4 Bayesian Models

4.1 Gaussian Mixture Model clustering and K-means clustering

The actual data of engine health is plotted Fig 8a for reference between the Engine RPM and Coolant Temperature based on the Engine condition. K means clustering is employed Fig 8b for these features and Gaussian Mixture Model (GMM) clustering is also employed Fig 8c. It can be seen that GMM is able to cluster the mixed data with a distribution unlike the K means clustering which is clustering based on the distance.



(a) Actual Data based on engine condition

(b) K-means clustering

(c) GMM clustering

Figure 2: Clustering Model comparison

4.2 Bayesian Classification

4.2.1 Using Bambi Library

Classification of Engine condition is done using Bambi Library[2] in python. The model specification is given as:
Family: bernoulli
Link: p = logit
Observations: 19535
Priors:
target = p
Common-level effects
Intercept: Normal(mu: 0.0, sigma: 71.313)
Fuel_pressure: Normal(mu: 0.0, sigma: 0.9055)
Coolant_pressure: Normal(mu: 0.0, sigma: 2.4123)
lub_oil_temp: Normal(mu: 0.0, sigma: 0.8036)
Coolant_temp: Normal(mu: 0.0, sigma: 0.4028)
Lub_oil_pressure: Normal(mu: 0.0, sigma: 2.4471)
Engine_rpm: Normal(mu: 0.0, sigma: 0.0093)

Classification results:

4.2.2 Using Bayesian ridge

Bayesian ridge regression is used for classification of engine health by rounding off the prediction to convert it into 1 or 0. Accuracy score: 0.087, Accuracy: 0.649

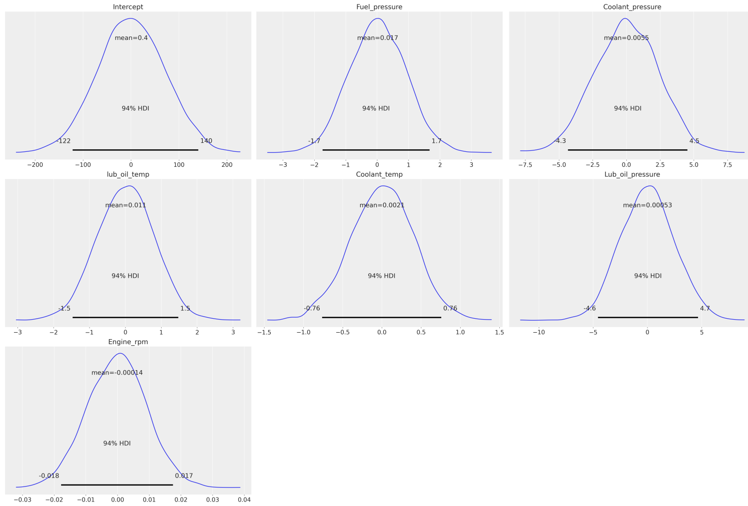


Figure 3: Prior Distribution

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
Intercept	93.08	91.535	-7.601	212.905	45.508	34.834	4	11	3.6
Fuel_pressure	0.061	0.594	-0.885	0.741	0.296	0.226	5	31	3.41
Coolant_pressure	-0.237	0.314	-0.768	0.036	0.156	0.12	4	11	3.63
lub_oil_temp	-0.02	0.608	-0.977	0.665	0.303	0.232	4	15	3.94
Coolant_temp	0.024	0.481	-0.394	0.819	0.24	0.184	4	11	3.98
Lub_oil_pressure	0.102	0.495	-0.69	0.66	0.247	0.189	4	13	3.25
Engine_rpm	-0.118	0.186	-0.338	0.134	0.092	0.071	4	11	4.06

Figure 4: Classification Results using Bambi

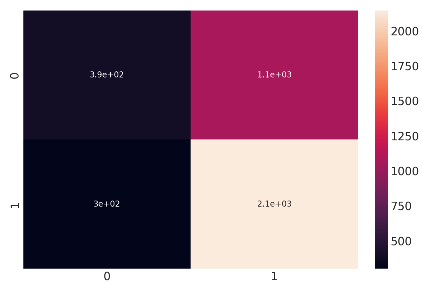


Figure 5: Confusion Matrix using Bayesian Ridge

4.3 Bayesian Linear Regression

4.3.1 Using Full PyMC model

Bayesian linear regression is done using full PyMC [3] model and bambi model. Regression is done for the Coolant Temperature by using all other features. The posterior distribution is given in Fig 6 The left side shows our marginal posterior – for each parameter value on the x-axis we get a probability on the y-axis that tells us how likely that parameter value is.

4.3.2 Using Bambi Model

Using Bambi library regression is done. Model is sampled using NUTS sampler. Model specifications are given as:
Family: gaussian
Link: mu = identity
Observations: 19535
Priors:

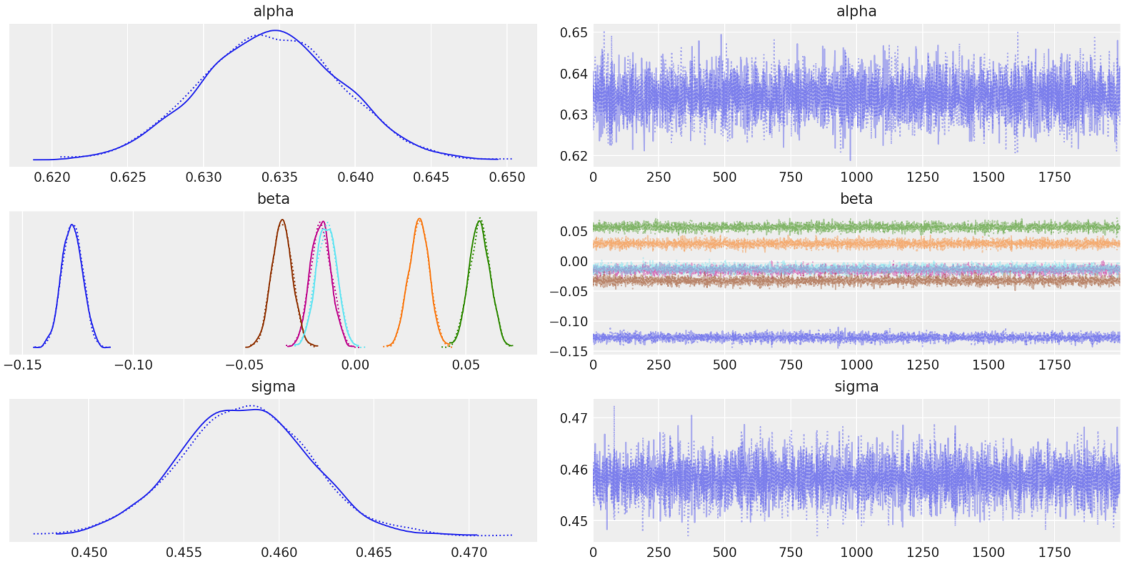


Figure 6: Posterior plots using PyMC

target = mu
Common-level effects
Intercept   Normal(mu: 78.4274, sigma: 397.3312)
Fuel_pressure   Normal(mu: 0.0, sigma: 5.62)
Coolant_pressure   Normal(mu: 0.0, sigma: 14.9722)
lub_oil_temp   Normal(mu: 0.0, sigma: 4.9878)
Lub_oil_pressure   Normal(mu: 0.0, sigma: 15.1882)
Engine_Condition   Normal(mu: 0.0, sigma: 32.1474)
Engine_rpm   Normal(mu: 0.0, sigma: 0.058)
Auxiliary parameters
sigma   HalfStudentT(nu: 4.0, sigma: 6.2066)

Priors plot are shown in Fig 7.

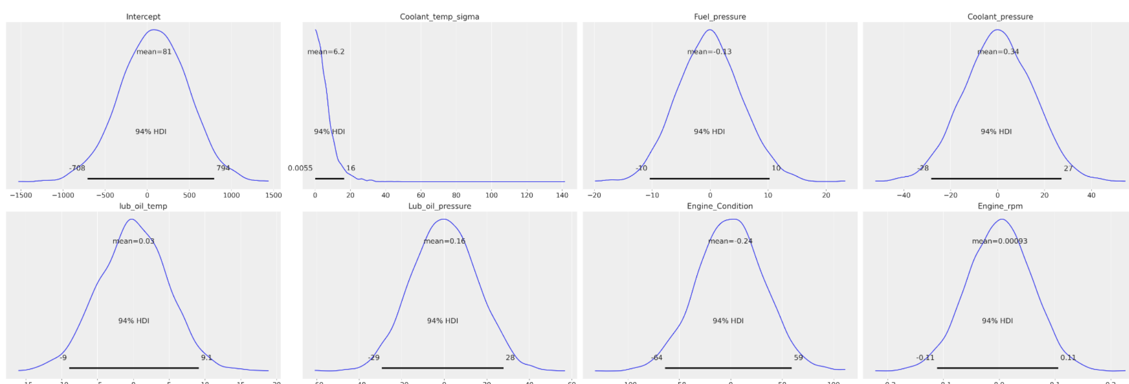
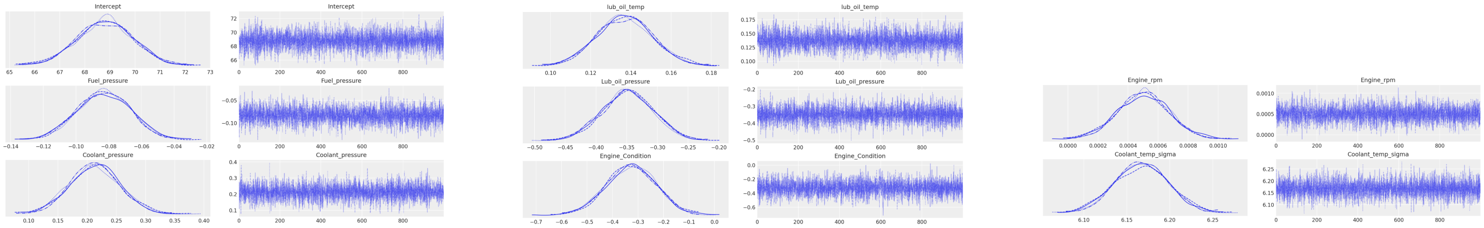


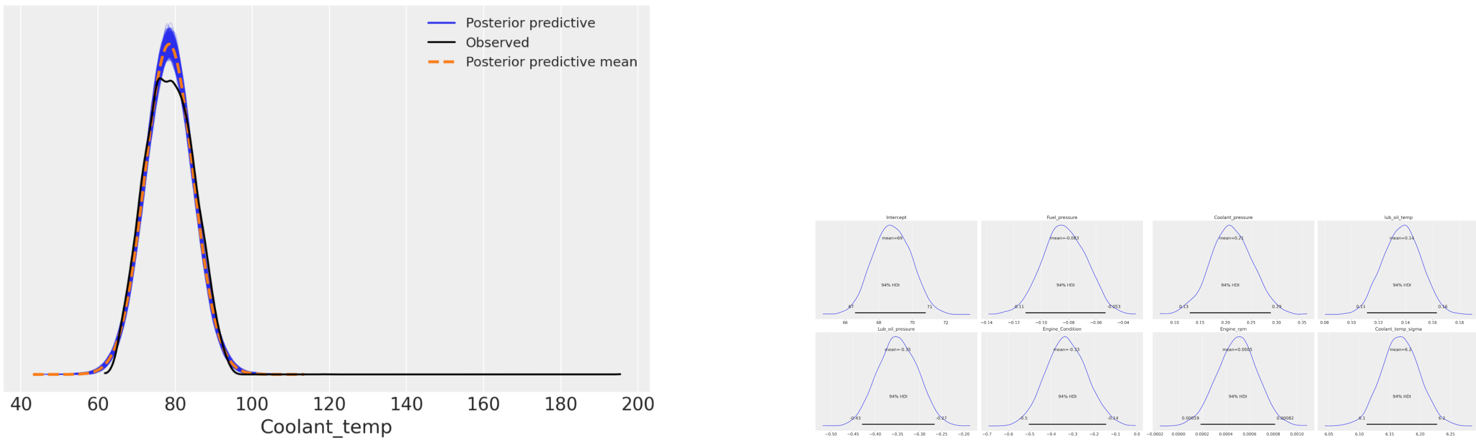
Figure 7: Priors using Bambi library

Results are shown in Fig 8 that is the diagnostic info on the model parameters.



(a) (b) (c)

Figure 8: Results of Linear regression using Bambi



(a) Posterior Prediction (b) Posterior Plots

Figure 9: Posterior prediction and posterior plots

5 Future Work

One potential use for this dataset could be to analyze the performance of different types of engines and vehicles. Researchers could use the data to compare the performance of engines from different manufacturers, for example, or to evaluate the effectiveness of different maintenance strategies. This could help drive innovation and improvements in the automotive industry.

Once the model is trained, it could be integrated into a larger system for monitoring the health of automotive engines. For example, sensors could be installed in vehicles to collect real-time data on engine performance, which is then sent to a central server for analysis. The predictive maintenance model could be run on this data, generating alerts or recommendations for maintenance or repair.

References

[1] Automotive Vehicles Engine Health Dataset. Kaggle;. https://www.kaggle.com/datasets/parvmodi/automotive-vehicles-engine-health-dataset/.

[2] Capretto T, Piho C, Kumar R, Westfall J, Yarkoni T, Martin OA. Bambi: A Simple Interface for Fitting Bayesian Linear Models in Python. Journal of Statistical Software. 2022;103(15):1–29. Available from: `https://www.jstatsoft.org/index.php/jss/article/view/v103i15`.

[3] Wiecki T, Salvatier J, Vieira R, Kochurov M, Patil A, Osthege M, et al.. pymc-devs/pymc: v5.12.0. Zenodo; 2024. Available from: `https://doi.org/10.5281/zenodo.10892107`.