1. Introduction

   Backpain is major concern for our society due to prolonged work hours. The requirement is to develop a model which is less compute intensive and provide good accuracy. The model is designed to deploy on a IOT device, hence the size of the model is critical for this project.

   Due to this requirement, we will be evaluating models which only saves weights the number of weights should be minimum.

2. Data where you describe the data that will be used to solve the problem and the source of the data.

   The source of data is from a Kaggle repository. The url is
   https://www.kaggle.com/sammy123/lower-back-pain-symptoms-dataset

   The data set contains following clinical measurements to predict if the back pain is normal or abnormal

   - The features are
   - pelvic_incidence
   - pelvic tilt
   - lumbar_lordosis_angle
   - sacral_slope
   - pelvic_radius
   - degree_spondylolisthesis
   - pelvic_slope
   - direct_tiltthoracic_slope
   - cervical_tilt
   - sacrum_angle
   - scoliosis_slope

3. Methodology

   Following exploratory analysis made.

   - There are no missing values.
   - The target prediction variable is imbalanced. Hence during the train test split stratify is used to include the same proportion in both train and test samples of target class ratio.
   - The box plot shows few readings as outliers, but according to domain expert the values are valid, hence did not remove the outlier readings.
   - The pair plot shows the correlation between the variables.
   - The values needs to be normalized and standardized. Using standardscalar the attributes brought to same scale.
   - Sklearn pipeline is used for easy deployment for new value prediction. The pipeline will takecare of transform of new attribute values during deployment using the same scale used for model development.

4. Results section where you discuss the results.

The result of

- svc model is about 81 % accuracy.
- Decision tree classifier is 77% accuracy.

5. Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
    - The number of samples are less, collecting more examples to train will improve the model accuracy without impacting the size of the model. (as we are using svc and decision tree classifier)
6. Conclusion section where you conclude the report.

    Since we have the requirement of size to deploy on end user device, the model size needs to be considered.

    - Size of svc model is 20k while size of decision tree classifier is only 7 k bytes.