**Relatório**

O algoritmo implementado cobre em parte as funções pedidas no enunciado. Implementa uma árvore através de nós com o correspondente antecessor e predecessor, guardando assim as informações necessárias para que seja possível trabalhar com os dados. Em cada nó está contida informação acerca da sua impureza, exemplos e atributo.

O algoritmo tem as funções fit() e predict() pedidas, sendo que recebe um ficheiro csv como base para que possa organizar os conjuntos de teste e treino para a realização do modelo. Para implementar o modelo é dada a classe para avaliar o mesmo construindo assim a tabela necessária ao cálculo da impureza dos diferentes atributos.

No cálculo da impureza dos atributos o utilizador pode escolher entre o índice de Gini ou Entropia para que seja utilizado. Durante a execução do método fit() é calculado a impureza dos atributos sendo que será guardada a informação de cada nó e os seus respetivos antecessores (por defeito será sempre usado o índice de Gini). Por fim é criado um output representativo da árvore em questão para que possa ser analisado com mais detalhe.

Na representação da árvore será tido em conta o método de pré poda caso este tenha sido devidamente inicializado na classe. No caso poderá ser definido um limite de folhas ou um limite de profundidade da árvore criada.

O algoritmo implementado não tem qualquer implementação de métodos pós poda.

O método predict(), analiza a árvore criada e compara com o conjunto para que possa ser calculada a exatidão do mesmo.

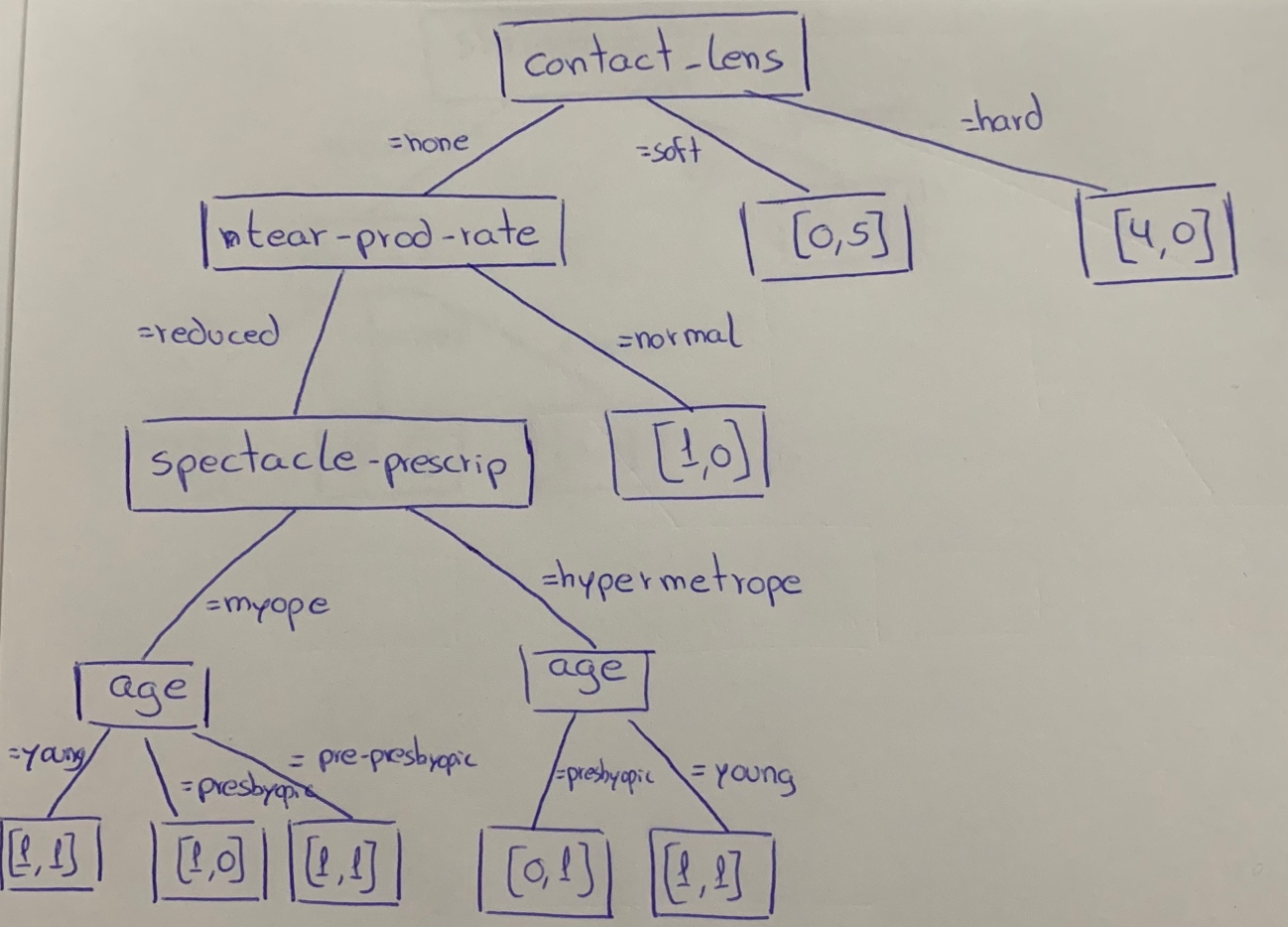
**Outputs**

Devido ao algoritmo desenvolvido apresentar uma forma de árvore que não é facilmente identificada, todos os outputs foram devidamente copiados para que fosse possível analisar as árvores obtidas sem que fosse necessário analisar exaustivamente o output do algoritmo.

**Contact-Lenses**

Os dados gerados ao analisar os dados do ficheiro “contact-lenses” são bastante precisos, obtendo uma precisão de 1 em todas as alterações feitas de pré poda nas árvores. Sendo que das três diferentes formas de construir árvores testadas ambas tinham precisão 1. Das árvores uma não tem qualquer método de poda associado, o segundo e o terceiro tipo têm um método de pré poda associado, sendo uma limitada em número de exemplos e a outra limitada em profundidade.

O esboço da árvore obtida pelo algoritmo sem qualquer método de poda seria o seguinte:



**Soybean**

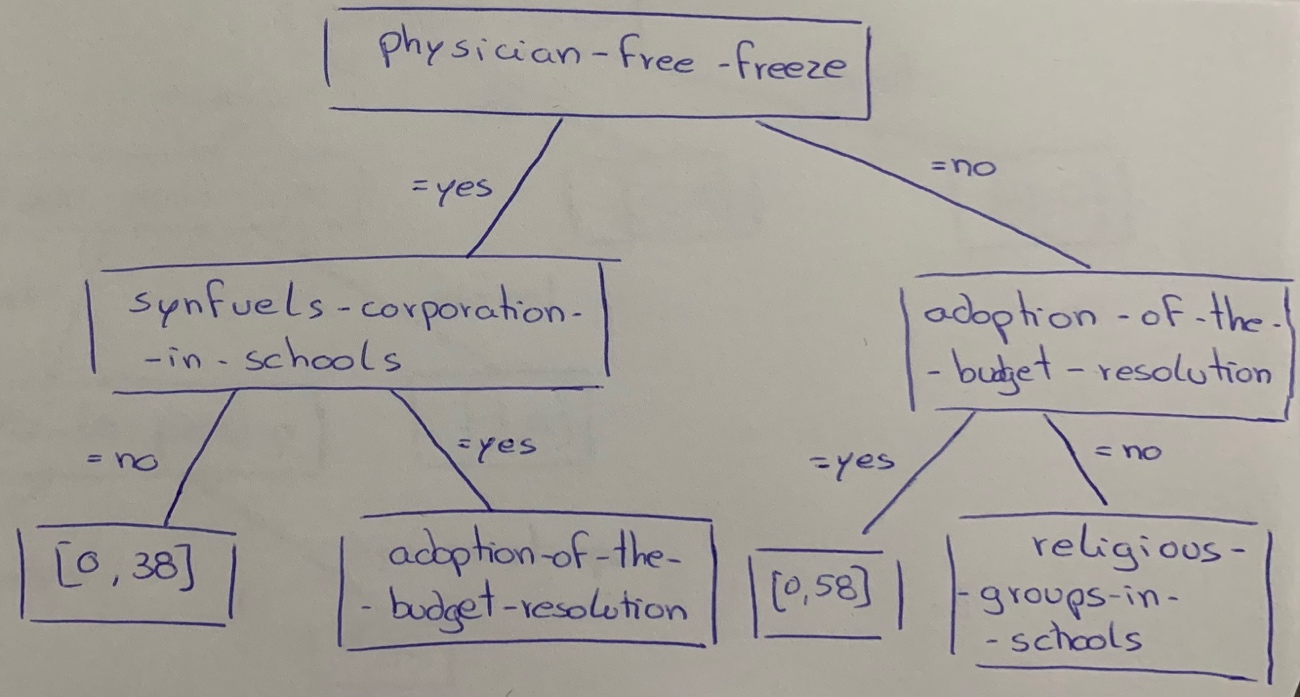
Para os dados de “soybean” foi utilizada a coluna “hail” como classe. Nesta experiência houve uma pequena alteração de precisão ao utilizar diferentes métodos de poda, ainda que nada que tivesse muito impacto nas árvores obtidas ou na sua precisão. Assim sendo na árvore obtida sem qualquer método de poda foi obtida uma precisão de 0.73. Nos dois modelos seguintes a precisão foi praticamente a mesma, sendo que ao limitar o número de exemplos a 5 a precisão obtida continua de 0.73, mas ao limitar o número de exemplos e a profundidade a precisão obtida desce para 0.72. Devido á extensão das árvores obtidas é apenas demonstrado um pequeno exemplo da última árvore com os dois métodos de pré poda:

Uma imagem com texto, quadro branco

Descrição gerada automaticamente

**Vote**

Na análise dos dados “vote”, continua a existir alguma semelhança na precisão assim como nos anteriores. Sendo que a árvore obtida sem qualquer tipo de poda demonstrou uma precisão de cerca de 0.57. De seguida são analisadas duas árvores com métodos de pré poda, sendo aplicado em ambos um limite de profundidade em ambas (de igual profundidade). Na primeira árvore construída apenas com o limite de profundidade foi obtida uma árvore com cerca de 0.57 de precisão.



Na árvore seguinte foi aplicada também um limite no número de exemplos por nó (20), sendo que a precisão aumentou ligeiramente para 0.59, o que pode demonstrar alguma inconsistência no número de exemplos entre os diferentes atributos.

Uma imagem com texto, quadro branco

Descrição gerada automaticamente