

#### Zadanie 4.2, 228884, Mateusz Guściora

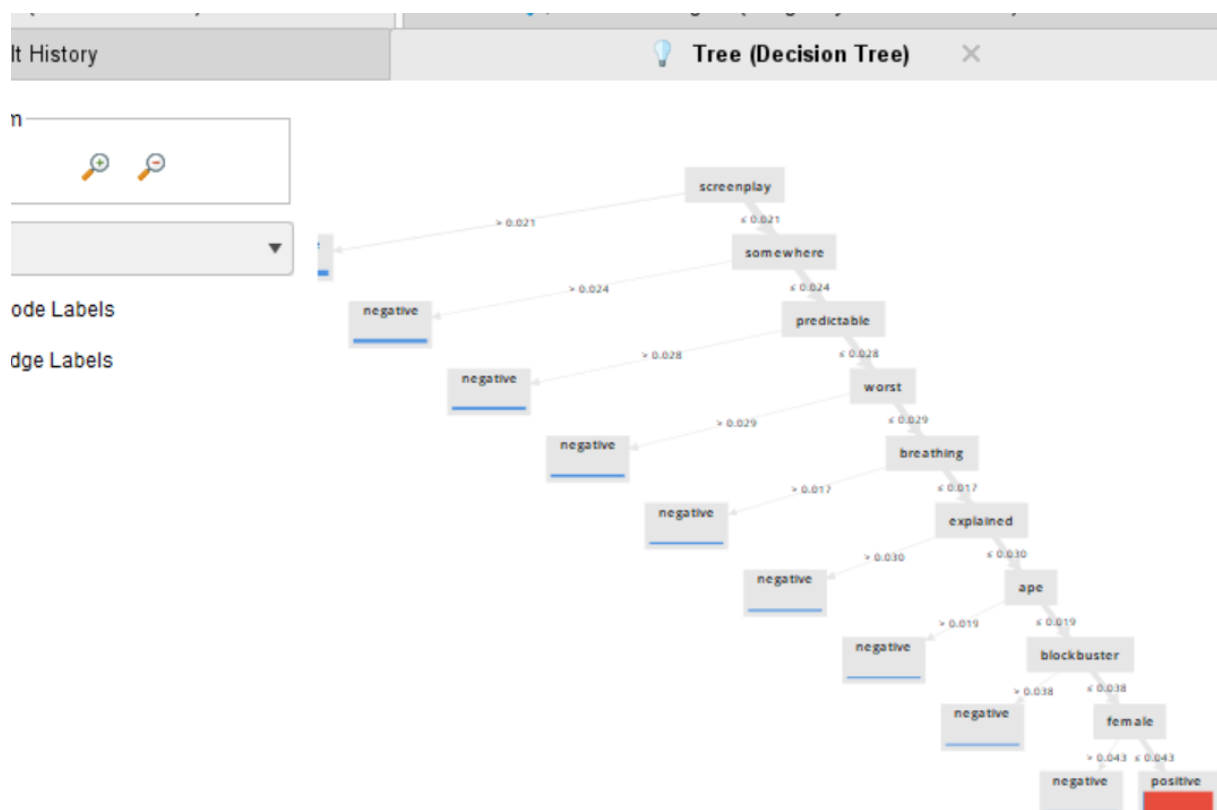
a) Analiza wydźwięku na przykładzie Sentiment Analysis.

W pierwszym podpunkcie uruchamiamy przykładowy proces analizy wydźwięku opinii. Dodajemy także operatory zapisu write as text i write as csv.

Accuracy 63% oraz kappa 0.244 to średnie wyniki predykcji.

b) Wstawiamy operator weight by information gain oraz Sample

c) W punkcie zmieniliśmy SMV na drzewo decyzyjne oraz dodaliśmy jako parametr w performance wykres AUC. Fragment budowy drzewa poniżej (również jest zapisany w plikach).



Operator weight by information gain wydłuża czas trwania procesu (o kilka minut)

d)

\*) Wstawienie operatora Extract Sentiment na wyjściu exa operatora Cross Validation. Należy po wstawieniu jeszcze wybrać text attribute, wybieramy z text (można jeszcze sentiment). Wpływa to nieznacznie na dokładność i wskaźnik kappa → oba obniżają swoje wartości

\*)\_2 Dodano operatory generate n grams do Process Documents (W strumieniu 1 i w strumieniu 2)

Dodanie parametru pozytywnie wpłynęło na jakość predykcji zwiększając accuracy oraz kappa

e) Analizując (powierzchownie) opinie dotyczyły filmów. Jako pozytywną klasę operator uznał np. „great film...” a jako negative np. („young enistein embarrassingly lame...”). Accuracy na poziomie 65% oraz kappa 0.244 co daje średnie wyniki.