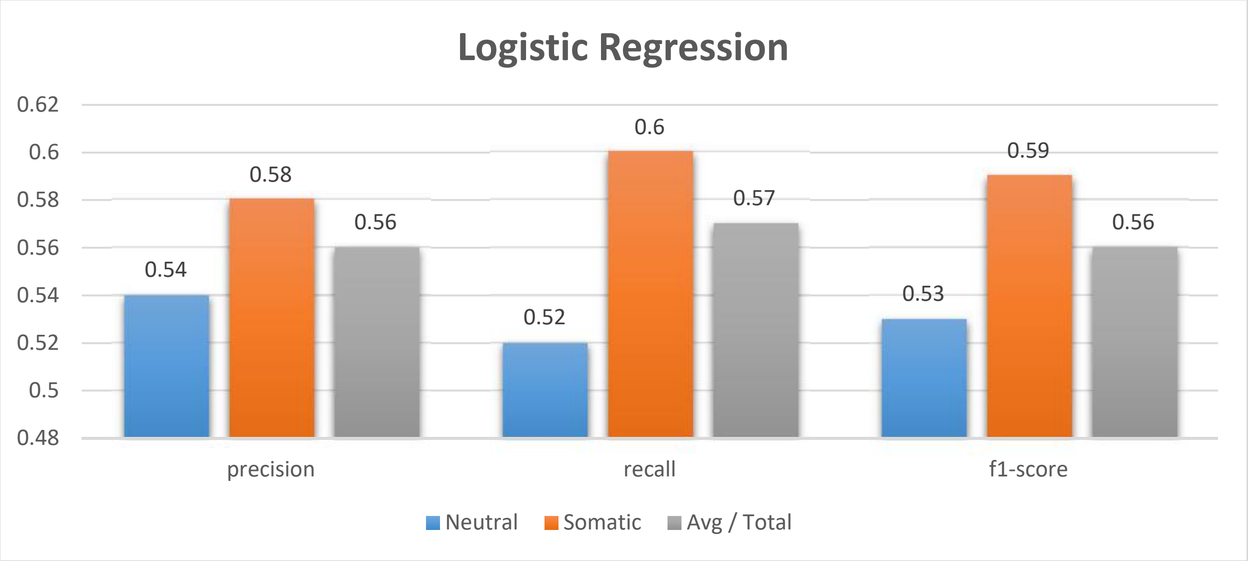
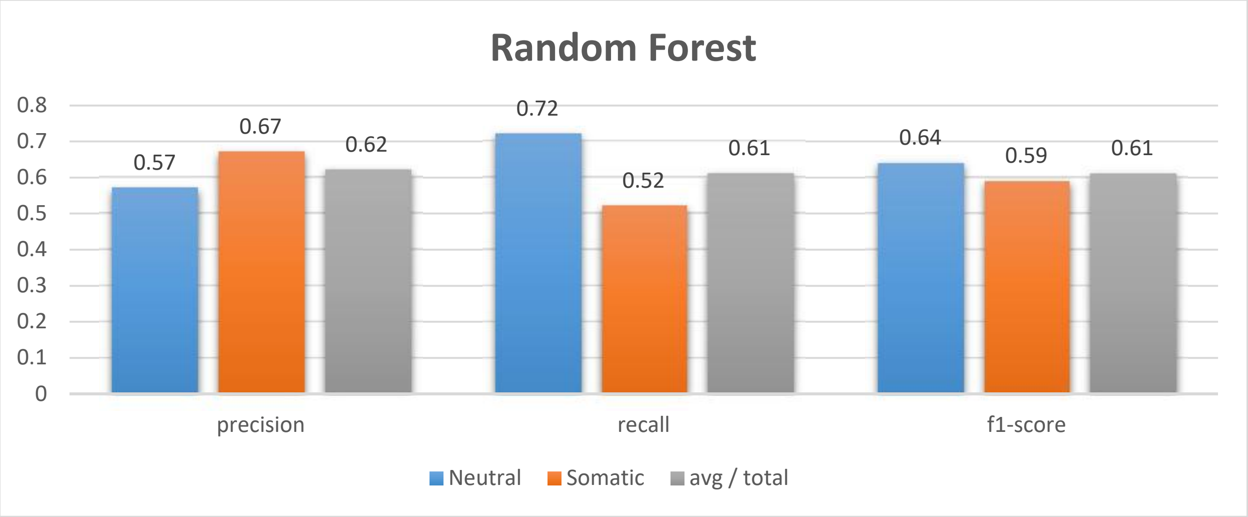
Inference from our project

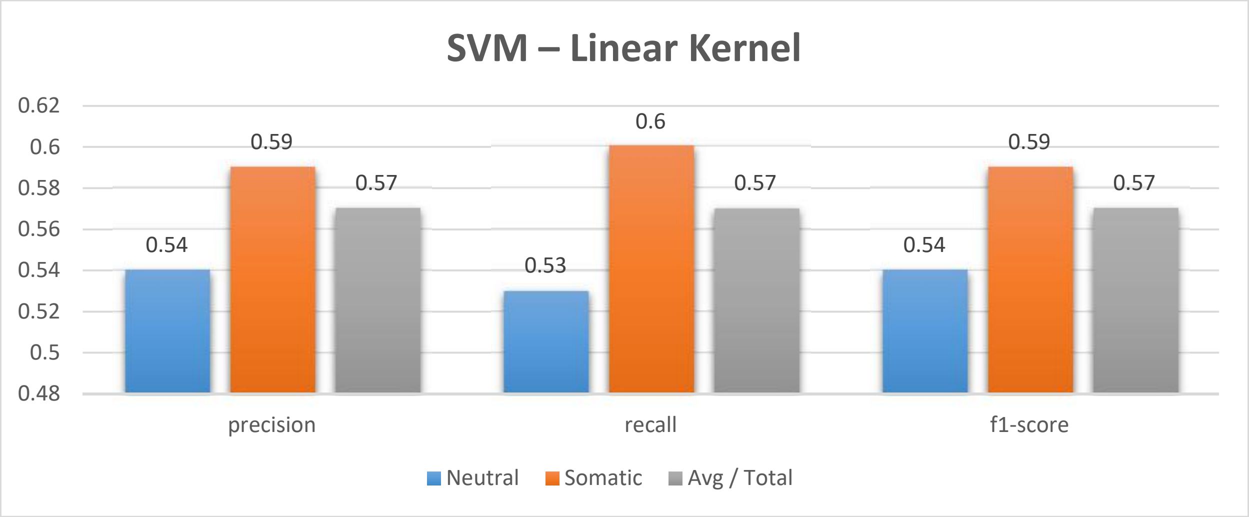
Logistic Regression

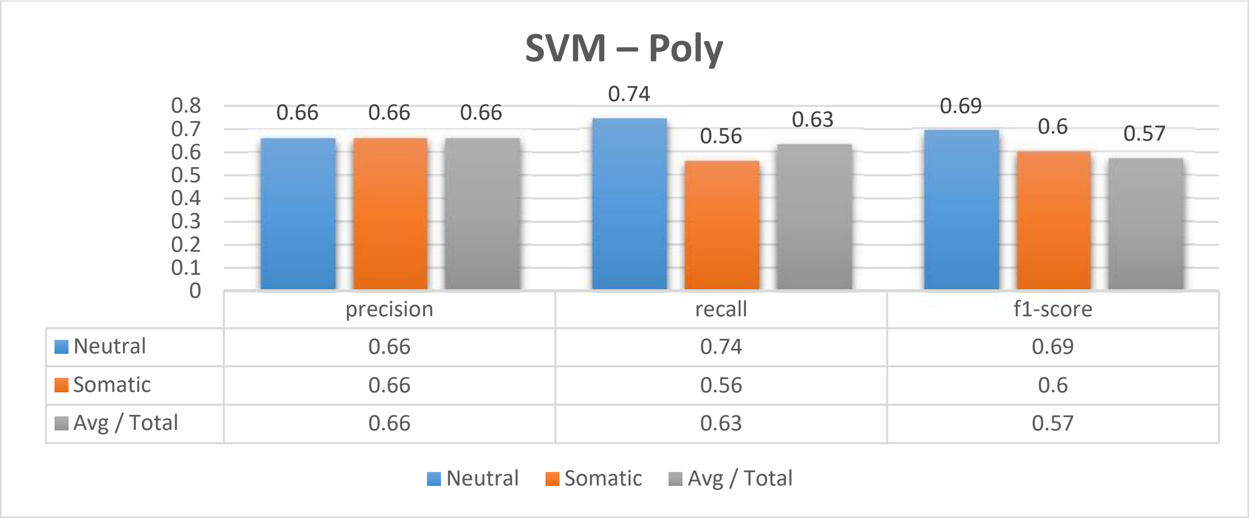


Random Forest



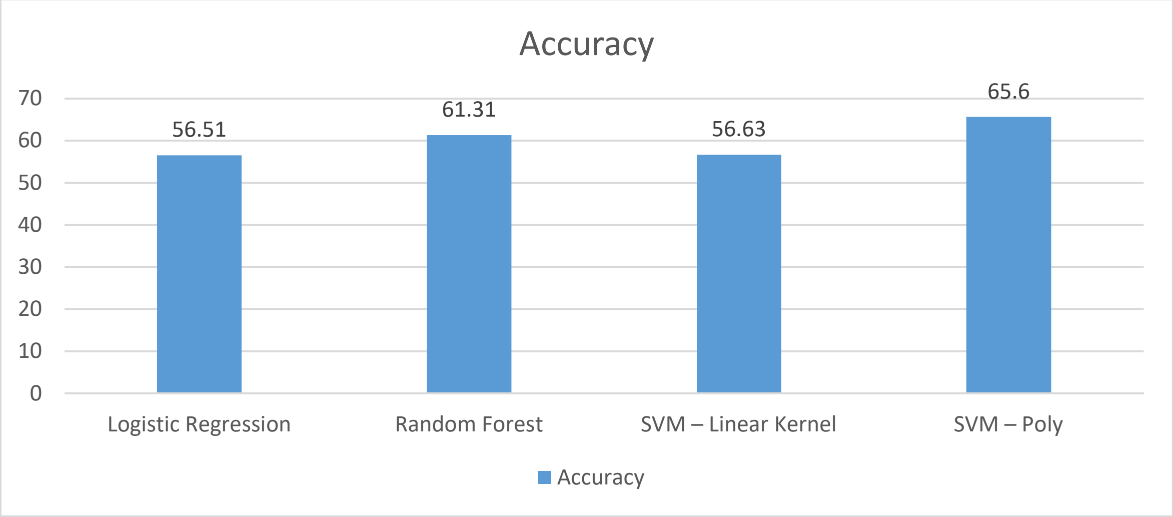
SVM –Linear Kernel


SVM – Linear Kernel

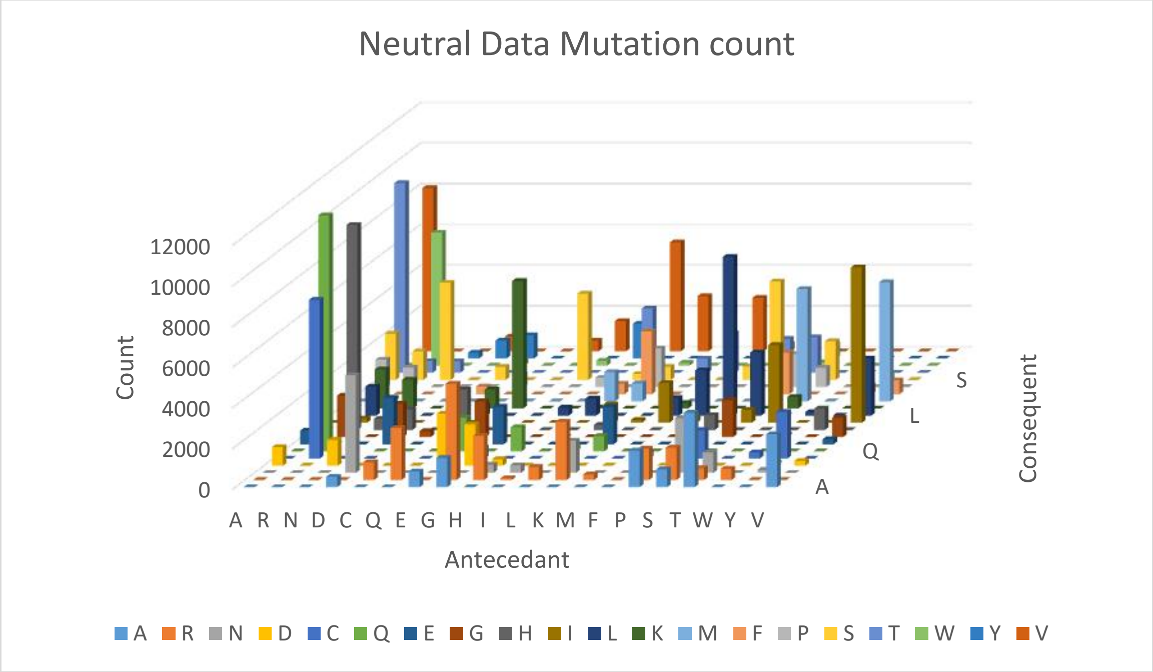

SVM – Poly

	precision	recall	f1-score
🟦 Neutral	0.66	0.74	0.69
🟧 Somatic	0.66	0.56	0.6
⬜ Avg / Total	0.66	0.63	0.57

Accuracy

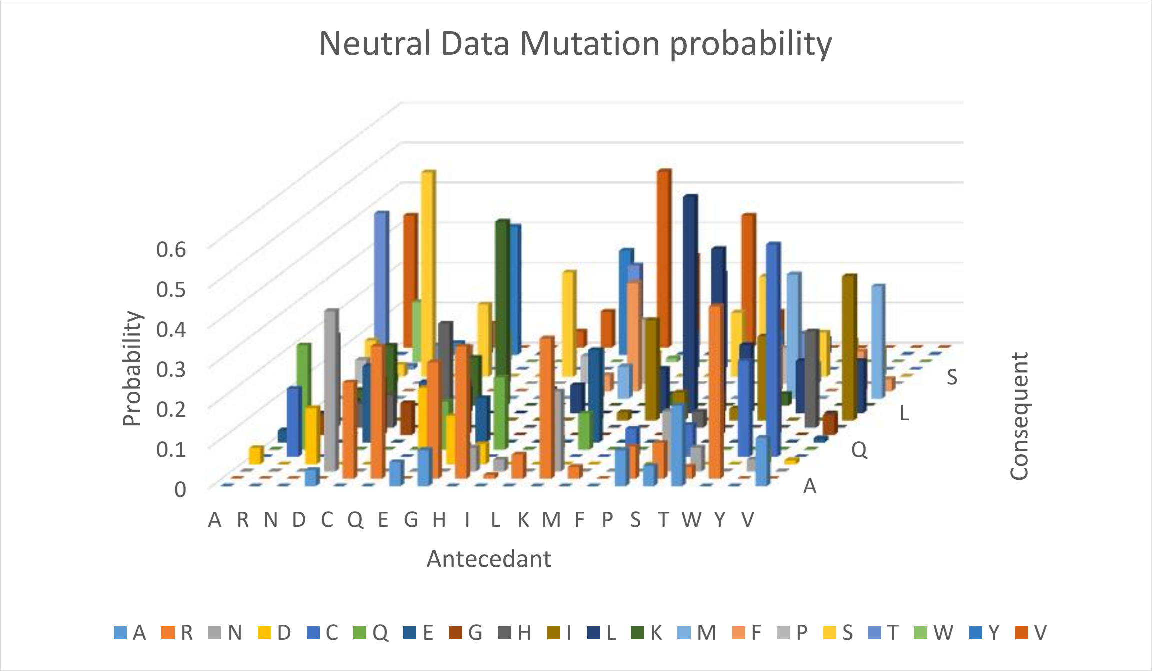


Neutral Data



Inference

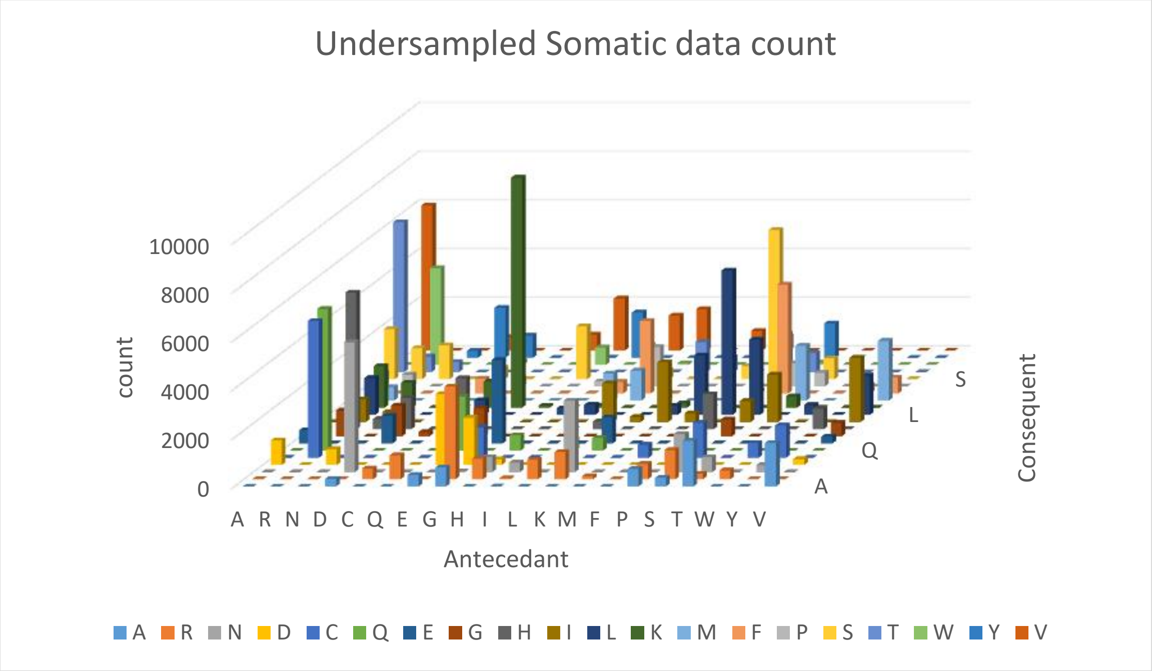
- The most frequently occurring mutation is R to Q and they occur 11601 times in the neutral dataset.
- The least occurring mutations are T to Q , S to D, P to D , D to S and Y to R and they occur only once in the neutral dataset



Inference

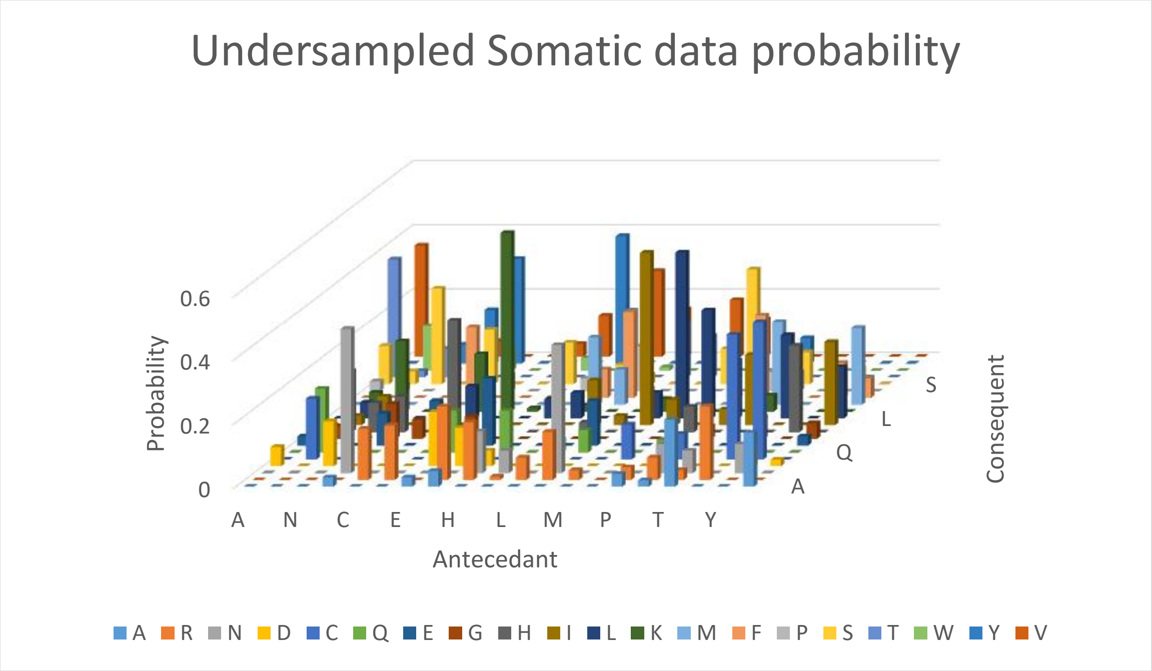
- The mutation F to L has the highest probability of occurrence 0.54.
- The mutations V to D , V to E , S to W, R to T, G to W, I to R, L to W, S to W have the lowest probability of occurrence of 0.01

Under Sampled data



Inference

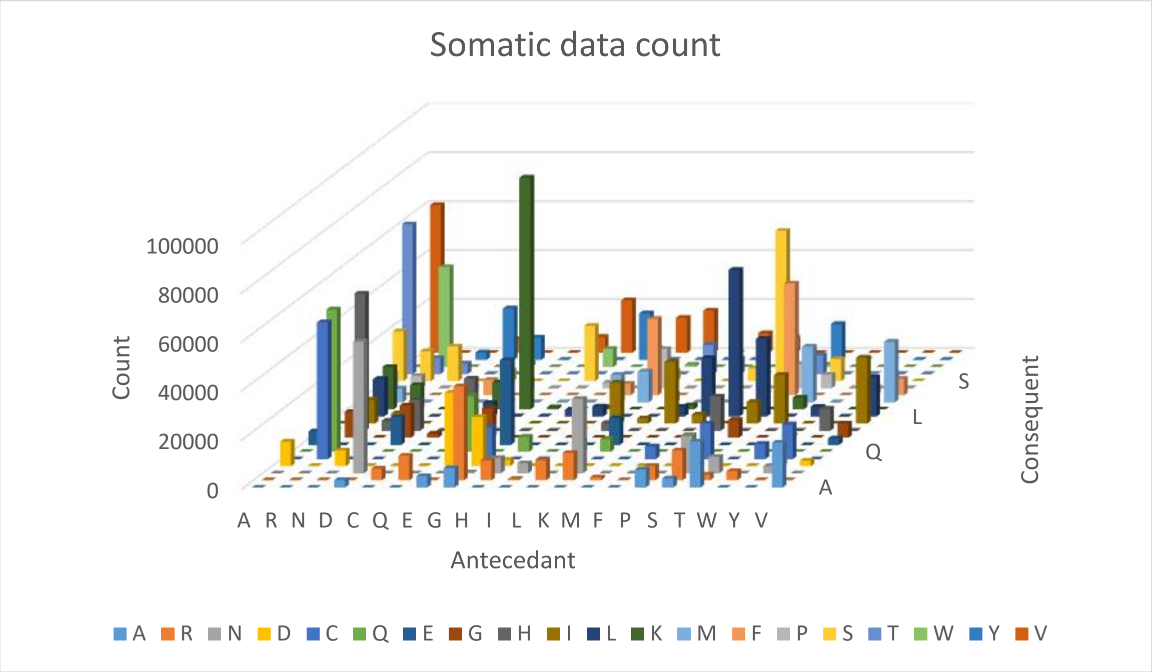
- The mutation E to K was the most frequently occurring mutation with an occurrence count of 9446
- There were numerous mutations with an occurrence count of one which indicated that there were lots of sparsely occurring mutations in the dataset



Inference

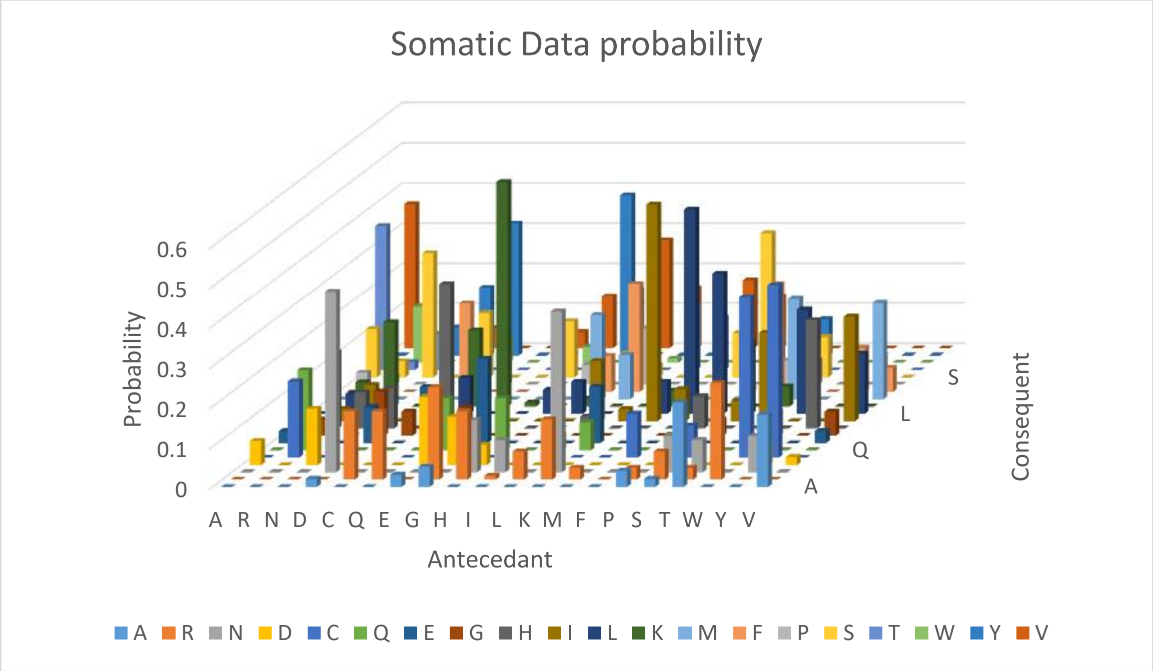
- The mutation E to K has the maximum probability of occurrence with a probability of occurrence of 0.56
- There were numerous mutations with the lowest probability of 0.01

Complete data



Inference

- The mutation E to K has the maximum count of occurrence with a count of 94674 times
- There were numerous mutations with an occurrence count of one which indicated that there were lots of sparsely occurring mutations in the dataset



Inference

- The mutation E to K has the maximum probability of occurrence with a probability of occurrence of 0.56
- There were numerous mutations with the lowest probability of 0.01.