

Použitie metód strojového učenia v oblasti biológie je často zložité. Na jednej strane sa v tejto oblasti zvyšuje počet premenných, ktoré je možné zaznamenať. Na strane druhej je však meranie pre každú entitu nákladné, a tak je počet pozorovaní často nízky. V tejto práci sa preto venujeme datasetom, ktoré majú nízky počet pozorovaní, ale veľký počet premenných. Zameriavame sa na rozličné kombinácie metód výberu premenných a klasifikačných metód a pokúšame sa odpovedať na otázku, ktoré kombinácie fungujú najlepšie. Pre lepšie pochopenie súvislostí medzi metódami pracujeme s dvoma simulačnými štúdiami a niekoľkými reálnymi datasetmi. Výsledky práce naznačujú, že väčšina klasifikačných metód dosahuje lepšie výsledky, ak pracujú s predvybranými premennými a nie so všetkými premennými dostupnými v dátach. V práci zároveň definujeme hranice pre počet pozorovaní datasetu, od ktorých dosahujú metódy výberu premenných vyššiu kvalitu a stabilitu. V závere práce popisujeme identifikovaný vzťah medzi mierou stability výberu premenných (tzv. Jaccardov index) a mierou kvality výberu premenných (tzv. False Discovery Rate).