Semi-supervised učení je technika strojového učení snažící se využít nejen označkovaná data (data pro která známe požadované výstupy), ale i neoznačkovaná data (data pro která požadované výstupy neznáme) s cílem snížit požadavky na množství označkovaných dat a tím umožnit použití strojového učení i v případech kdy je označkování velkého množství dat příliš náročné. I přes svůj rychlý vývoj v posledních letech stále trpí problémy které brání jeho širokému využití v praxi. Jedním z těchto problémů je nesoulad distribucí tříd. Ten vzniká, když neoznačkovaná data obsahují vzorky které nepatří do žádné ze tříd označkovaných dat. To může zmást učení klasifikátoru do takové míry, že je ve výsledku horší než kdyby neoznačkovaná data vůbec nebyla využita.
Tato diplomová práce navrhuje metodu nazvanou Unfavorable Data Filtering (UDF), která nejprve z dat extrahuje důležité příznaky a pak se na jejich základě pomocí filtru založeného na podobnosti datových vzorků snažít vyřadit nerelevantní data z trénovacích dat. Díky tomu, že je UDF použita před semi-supervised učením je možné ji použít s libovolnou učící metodou. Pro zjištění jak efektivní UDF je jsme provedli mnoho experimentů, převážně na datasetu zvaném CIFAR-10. Pomocí těchto experimentů jsme zjistili, že filtrování pomocí UDF je opravdu schopno výrazně vylepšit výsledky učeného klasifikátoru, identifikovali základní zásady kdy se ho vyplatí použít a objevili důležitou vlastnost nesouladu distribucí tříd.