Cílem práce je vyhodnotit klasifikaci českého textu s malým množstvím trénovacích dat. Používáme tři datasety, z nichž dva jsou veřejně dostupné a jeden je vytvořen částečně námi. Základ tohoto datasetu tvoří smlouvy, které nám poskytla webová platforma Hlídač Státu. Většina dat je klasifikovaná automaticky a jen malá část ručně. Jeho charakteristickým znakem je, že obsahuje dlouhé smlouvy v českém jazyce. S navrženým modelem dosahujeme na veřejně dostupných datasetech velmi dobrých výsledků, což potvrzuje dostatečný výkon našeho modelu. Navíc jsme na těchto veřejně dostupných datasetech provedli experimentální měření zašuměných dat a různého množství dat potřebných k natrénování modelu. Na datasetu smluv jsme se zaměřili na výběr správné části z jednotlivých smluv a zkoumali jsme, pomocí které části můžeme dosáhnout nejlepší výsledků. Zjistili jsme, že u datasetu, který z důvodu automatického anotování obsahuje jistou část systematických chyb, je pro klasifikaci výhodnější použít kratší, ale relevantnější část smlouvy, než vzít ze smlouvy delší text a spoléhat se, že BERT se z toho naučí správně.