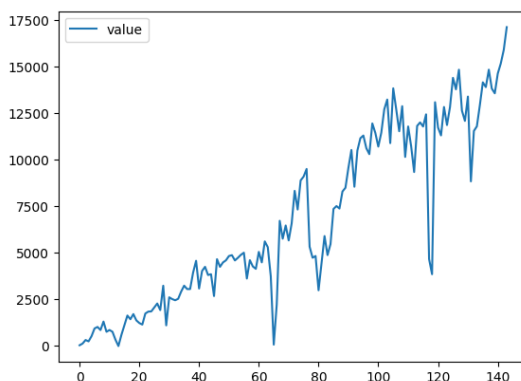


Predikce - vstupní úkol

Poskytnutá data představují časovou řadu, která má určitý trend se značným zašuměním.

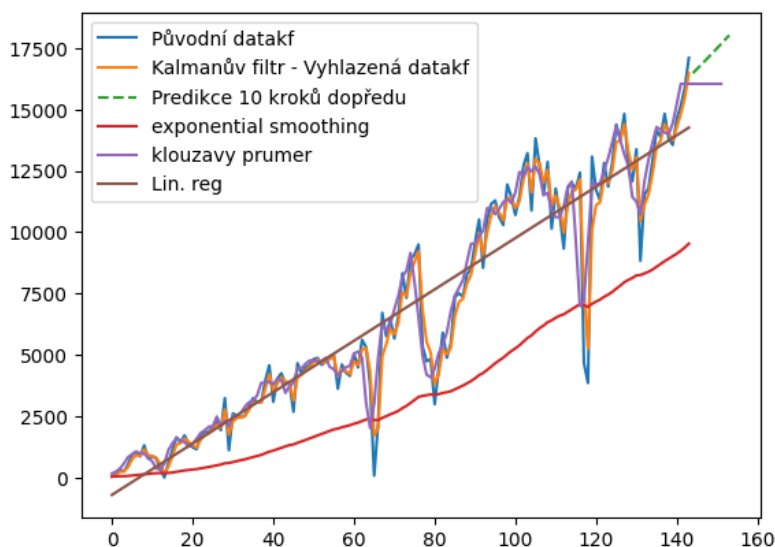


Tato skutečnost značně komplikuje úlohu predikce, neb lehčí modely budou špatně odhadovat vývoj dat. Mezi lehčí modely, které by se na úlohu extrapolace (predikce neznámých budoucích dat) hodily se řadí klouzavý průměr, autoregrese, exponenciální vyhlazování, či lineární regrese.

Problém s nasazením klouzavého průměru je, že data pouze vyhlazuje, a tedy neukáže očekávaný směr do budoucnosti. Problém s exponenciálním vyhlazováním je jeho náchylnost na nastavení hyperparametru alfa a nestabilní data. Problém u autoregrese jsou nestacionární data a nutná znalost, jak stará data mají vliv na aktuální stav. A nakonec problém s lineární regresí je ten, že v tomto případě zachycuje pouze přibližný trend vývoje časové řady a nebere v potaz možný vliv jen části minulých dat. V případě sezónosti se bude chyba modelu daleko více projevovat.

Mezi složitější modely, které by bylo vhodné uvážit pro modelování budoucích prodejů lze uvážit (S)ARIMA nebo Kalmanův filtr. ARIMA kombinuje sílu jak autoregresního modelu, klouzavého průměru a integrovaného modelu. Kalmanův filtr lépe zastihuje stav dynamického systému z měření, kde mohl hrát roli i šum nebo nejistota. U kalmánova filtru by se vyplatil kontinuální update každý týden, pro přesnější predikce. Sice odhadne do deseti týdnů dopředu, ale nejistota bude stoupat s každým krokem/týdnem.

Pro představu jsem provedl několik experimentů k vymodelování dat a budoucí predikce.



Pro tento úkol jsem se rozhodl pro vytvoření vlastní jednoduché lineární regrese bez hyperparametrů.