Definice problému:

Našim cílem je předpovídat cenu ojetých aut na základě atributů uvedených v inzerátech na webové stránce [ebay-kleinanzeigen.de](https://www.ebay-kleinanzeigen.de/).

*Jaký je očekáváný výstup?*

Našim cílem je vytvořit a naučit model, který na základě daných atributů bude schopen predikovat očekávánou cenu auta. Jedná se tedy o úlohu regrese, proto vyzkoušíme přístupy typické pro tuto úlohu:

* Lineární regresi
* Lasso regresi

*Jaké algoritmy použijeme?*

## Popis použitých dat

Pro tento projekt jsme použili dataset s atributy ojetých aut, které byly staženy z inzerátu z Ebay-Kleinanzeigen (Leka, 2020). Těmito atributy jsou:

* dateCrawled: datum stažení inzerátu
* name: jméno inzerátu
* seller: prodejce (osobní nebo dealer)
* offerType: typ inzerátu (nabízející nebo poptávající)
* **price: cena auta v eurech**
* abtest:
* vehicleType: typ vozidla (např. suv, cabrio)
* yearOfRegistration: rok registrace
* gearbox: typ řazení (manual nebo automatický)
* powerPS : výkon auta
* model: model auta
* kilometer : počet najetých kilometrů
* monthOfRegistration : v jakém měsíci byl inzerát publikován
* fuelType: typ paliva
* brand: značka auta
* notRepairedDamage : zda má auto neopravenou závadu
* dateCreated : datum vytvoření inzerátu
* nrOfPictures : počet fotek auta
* postalCode: poštovní směrovací číslo
* lastSeenOnline : čas, kdy byl inzerát naposledy aktuální

Z těchto 19 atributů potřebujeme nalézt ty, které mají největší vliv na cenu a které následně použijeme jako features pro náš model

## Reference

Leka, O., 2020. *Used Cars Database*. [online] Kaggle.com. Available at: <https://www.kaggle.com/orgesleka/used-cars-database> [Accessed 13 June 2020].

Features: