# Strojové učení

**Strojové učení** je podoblastí [umělé inteligence](https://cs.wikipedia.org/wiki/Um%C4%9Bl%C3%A1_inteligence), zabývající se [algoritmy](https://cs.wikipedia.org/wiki/Algoritmus) a technikami, které umožňují počítačovému systému 'učit se'. Učením v daném kontextu rozumíme takovou změnu vnitřního stavu systému, která zefektivní schopnost přizpůsobení se změnám okolního prostředí.

Strojové učení se značně prolíná s oblastmi [statistiky](https://cs.wikipedia.org/wiki/Statistika) a [dobývaní znalostí](https://cs.wikipedia.org/wiki/Data_mining) a má široké uplatnění. Jeho techniky se využívají např. v [biomedicínské informatice](https://cs.wikipedia.org/w/index.php?title=Biomedic%C3%ADnsk%C3%A1_informatika&action=edit&redlink=1) (tzv. [systémy pro podporu rozhodování](https://cs.wikipedia.org/wiki/Syst%C3%A9my_pro_podporu_rozhodov%C3%A1n%C3%AD)), rozlišení nelegálního užití kreditních karet, [rozpoznávání řeči](https://cs.wikipedia.org/wiki/Rozpozn%C3%A1v%C3%A1n%C3%AD_%C5%99e%C4%8Di) a psaného textu, či mnohé další. Algoritmy však mohou být tendenční.[[1]](https://cs.wikipedia.org/wiki/Strojov%C3%A9_u%C4%8Den%C3%AD#cite_note-1)

## Obsah

* [1 Základní rozdělení algoritmů učení](https://cs.wikipedia.org/wiki/Strojov%C3%A9_u%C4%8Den%C3%AD#Základní_rozdělení_algoritmů_učení)
* [2 Základní druhy úloh](https://cs.wikipedia.org/wiki/Strojov%C3%A9_u%C4%8Den%C3%AD#Základní_druhy_úloh)
* [3 Podoblasti strojového učení](https://cs.wikipedia.org/wiki/Strojov%C3%A9_u%C4%8Den%C3%AD#Podoblasti_strojového_učení)
* [4 Terminologie](https://cs.wikipedia.org/wiki/Strojov%C3%A9_u%C4%8Den%C3%AD#Terminologie)
* [5 Software](https://cs.wikipedia.org/wiki/Strojov%C3%A9_u%C4%8Den%C3%AD#Software)
* [6 Reference](https://cs.wikipedia.org/wiki/Strojov%C3%A9_u%C4%8Den%C3%AD#Reference)
* [7 Související články](https://cs.wikipedia.org/wiki/Strojov%C3%A9_u%C4%8Den%C3%AD#Související_články)

## Základní rozdělení algoritmů učení

[Algoritmy](https://cs.wikipedia.org/wiki/Algoritmus) strojového učení lze podle způsobu učení rozdělit do následujících kategorií:

* [učení s učitelem](https://cs.wikipedia.org/wiki/U%C4%8Den%C3%AD_s_u%C4%8Ditelem) ([en:supervised learning](https://en.wikipedia.org/wiki/supervised_learning" \o "en:supervised learning)) Pro vstupní data je určen správný výstup (třída pro klasifikaci nebo hodnota pro regresi)
* [učení bez učitele](https://cs.wikipedia.org/w/index.php?title=U%C4%8Den%C3%AD_bez_u%C4%8Ditele&action=edit&redlink=1) ([en:unsupervised learning](https://en.wikipedia.org/wiki/unsupervised_learning" \o "en:unsupervised learning)) Ke vstupním datům není známý výstup
* kombinace učení s učitelem a bez učitele ([en:semi-supervised learning](https://en.wikipedia.org/wiki/semi-supervised_learning" \o "en:semi-supervised learning)) Část vstupních dat je se známým výstupem, ale další data, typicky větší, jsou bez něj. Často se používá [EM algoritmus](https://cs.wikipedia.org/w/index.php?title=Expectation%E2%80%93maximization_algoritmus&action=edit&redlink=1) ([en:Expectation–maximization algorithm](https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm" \o "en:Expectation–maximization algorithm)). Podobný přístup je [transdukce](https://cs.wikipedia.org/w/index.php?title=Transdukce_(strojov%C3%A9_u%C4%8Den%C3%AD)&action=edit&redlink=1)
* [zpětnovazebné učení](https://cs.wikipedia.org/wiki/Zp%C4%9Btnovazebn%C3%AD_u%C4%8Den%C3%AD) ([en:reinforcement learning](https://en.wikipedia.org/wiki/reinforcement_learning" \o "en:reinforcement learning)), též *učení posilováním*

Podle způsobu zpracování lze algoritmy rozdělit na

* dávkové: Všechna data požadují před začátkem výpočtu.
* inkrementální: Dokážou se "přiučit", tj. upravit model, pokud dostanou nová data, bez přepočítání celého modelu od začátku

## Základní druhy úloh

* [Klasifikace](https://cs.wikipedia.org/wiki/Klasifikace_(um%C4%9Bl%C3%A1_inteligence)) rozděluje vstupní data do dvou nebo několika tříd.
* [Regrese](https://cs.wikipedia.org/wiki/Regrese) odhaduje číselnou hodnotu výstupu podle vstupu
* [Shlukování](https://cs.wikipedia.org/wiki/Shlukov%C3%A1_anal%C3%BDza) zařazuje objekty do skupin s podobnými vlastnostmi, typicky při učení bez učitele

Další typy úloh jsou:

* [Ranking](https://cs.wikipedia.org/w/index.php?title=Rankink_(strojov%C3%A9_u%C4%8Den%C3%AD)&action=edit&redlink=1) určuje pořadí datových bodů, výsledkem je částečné nebo úplné setřídění
* Učení strukturovaných dat. *Výstupní* neboli hledaná struktura může být například sekvence, strom, graf, matice ... Aplikace jsou např. učení syntaktických *stromů* ve [zpracování přirozeného jazyka](https://cs.wikipedia.org/wiki/Zpracov%C3%A1n%C3%AD_p%C5%99irozen%C3%A9ho_jazyka), zarovnání *několika sekvencí* proteinů v [bioinformatice](https://cs.wikipedia.org/wiki/Bioinformatika), převod [řeči na textový](https://cs.wikipedia.org/wiki/Rozpozn%C3%A1v%C3%A1n%C3%AD_%C5%99e%C4%8Di) řetězec, tj. na *sekvenci* znaků, hledání vhodné molekuly reprezentované jako *graf* v [chemoinformatice](https://cs.wikipedia.org/wiki/Chemoinformatika" \o "Chemoinformatika), výstup [obrázkového algoritmu](https://cs.wikipedia.org/wiki/Zpracov%C3%A1n%C3%AD_obrazu) jako *matice* (mnoho konkrétních úloh) ...

## Podoblasti strojového učení

* Používané modely:
  + [Rozhodovací stromy](https://cs.wikipedia.org/wiki/Rozhodovac%C3%AD_stromy)
  + [Algoritmus k-nejbližších sousedů](https://cs.wikipedia.org/wiki/Algoritmus_k-nejbli%C5%BE%C5%A1%C3%ADch_soused%C5%AF)
  + Podpůrné vektory, viz [Support vector machines](https://cs.wikipedia.org/wiki/Support_vector_machines)
  + [Lineární diskriminační analýza](https://cs.wikipedia.org/w/index.php?title=Line%C3%A1rn%C3%AD_diskrimina%C4%8Dn%C3%AD_anal%C3%BDza&action=edit&redlink=1) ([en:Linear discriminant analysis](https://en.wikipedia.org/wiki/Linear_discriminant_analysis" \o "en:Linear discriminant analysis))
  + [Kvadratická diskriminační analýza](https://cs.wikipedia.org/w/index.php?title=Kvadratick%C3%A1_diskrimina%C4%8Dn%C3%AD_anal%C3%BDza&action=edit&redlink=1) ([en:Quadratic discriminant analysis](https://en.wikipedia.org/wiki/Quadratic_discriminant_analysis" \o "en:Quadratic discriminant analysis))
  + [Množina rozhodovacích pravidel](https://cs.wikipedia.org/w/index.php?title=Mno%C5%BEina_rozhodovac%C3%ADch_pravidel&action=edit&redlink=1)
  + [Perceptron](https://cs.wikipedia.org/wiki/Perceptron)
  + [Bayesovské sítě](https://cs.wikipedia.org/wiki/Bayesovsk%C3%A9_s%C3%ADt%C4%9B)
  + [Neuronové sítě](https://cs.wikipedia.org/wiki/Neuronov%C3%A9_s%C3%ADt%C4%9B)
* Techniky pro kombinaci více modelů ([en:Ensemble learning](https://en.wikipedia.org/wiki/Ensemble_learning" \o "en:Ensemble learning) a [en:Meta learning (computer science)](https://en.wikipedia.org/wiki/Meta_learning_(computer_science)))
  + [Bootstrap aggregating](https://cs.wikipedia.org/w/index.php?title=Bootstrap_aggregating&action=edit&redlink=1) (resp. zkratka [Bagging](https://cs.wikipedia.org/w/index.php?title=Bagging&action=edit&redlink=1)) ([en:Bootstrap aggregating](https://en.wikipedia.org/wiki/Bootstrap_aggregating" \o "en:Bootstrap aggregating))
  + [Boosting](https://cs.wikipedia.org/w/index.php?title=Boosting&action=edit&redlink=1) ([en:Boosting](https://en.wikipedia.org/wiki/Boosting" \o "en:Boosting))
  + [Stacking](https://cs.wikipedia.org/w/index.php?title=Stacking&action=edit&redlink=1)
* Testování přesnosti modelu:
  + [Křížová validace](https://cs.wikipedia.org/wiki/K%C5%99%C3%AD%C5%BEov%C3%A1_validace)
  + [Bootstrap (statistika)](https://cs.wikipedia.org/w/index.php?title=Bootstrap_(statistika)&action=edit&redlink=1) ([en:Bootstrapping (statistics)](https://en.wikipedia.org/wiki/Bootstrapping_(statistics)" \o "en:Bootstrapping (statistics)))
  + viz také odstavec [Odhady pravděpodobností správné klasifikace](https://cs.wikipedia.org/wiki/Diskrimina%C4%8Dn%C3%AD_anal%C3%BDza#Odhady_pravděpodobností_správné_klasifikace)

## Terminologie

* Data, body, případy, měření
* Atributy, rysy, proměnné, fíčury/features
* Druhy/typy atributů: binární, kategoriální (např. "A", "B", "AB" nebo "O" pro krevní skupiny, ordinální (např. "velký", "střední" nebo "malý"), celočíselné (např. počet výskytů slova v emailu) anebo reálné (např. měření krevního tlaku); strukturované, hierarchické

## Software

[RapidMiner](https://cs.wikipedia.org/w/index.php?title=RapidMiner&action=edit&redlink=1), [KNIME](https://cs.wikipedia.org/w/index.php?title=KNIME&action=edit&redlink=1), [Weka](https://cs.wikipedia.org/wiki/Weka), [ODM](https://cs.wikipedia.org/w/index.php?title=ODM&action=edit&redlink=1), [Shogun toolbox](https://cs.wikipedia.org/w/index.php?title=Shogun_toolbox&action=edit&redlink=1), [Orange](https://cs.wikipedia.org/w/index.php?title=Orange_(software)&action=edit&redlink=1), [Apache Mahout](https://cs.wikipedia.org/w/index.php?title=Apache_Mahout&action=edit&redlink=1) a [scikit-learn](https://cs.wikipedia.org/w/index.php?title=Scikit-learn&action=edit&redlink=1) jsou [softwarové balíky](https://cs.wikipedia.org/w/index.php?title=Softwarov%C3%BD_bal%C3%ADk&action=edit&redlink=1), které obsahují různé algoritmy strojového učení.

Online: Microsoft Azure Machine Learning (Azure ML)

## Reference

 <https://techxplore.com/news/2017-07-tackle-bias-algorithms.html> – Researchers tackle bias in algorithms