



# **Efektivní časoprostorové predikce s využitím metod strojového učení**

(ESPRED)

## **Metodika pro testování a validaci časoprostorových predikcí komfortu osob**

Projekt	Efektivní časoprostorové predikce s využitím metod strojového učení
Kód projektu	TJ04000232
Zhotovitel	Vysoké učení technické v Brně
Odpovědný autor	Ing. Martina Pálková (rozená Kratochvílová)
Další autoři	Ing. Uhlík Ondřej Ing. Okřinová Petra Ing. Vacková Jana Ing. Apeltauer Jiří, Ph.D. doc. Mgr. Apeltauer Tomáš, Ph.D. doc. Dr.techn. Ing. Podroužek Jan
Datum zhotovení	30. 05. 2022

## OBSAH

SEZNAM POUŽITÝCH ZKRATEK .....	4
TERMÍNY A DEFINICE .....	4
1 ÚVOD .....	5
2 ZÁKLADNÍ SROVNÁVACÍ ÚLOHY A JEJICH PARAMETRIZACE .....	7
2.1 TVORBA ZÁKLADNÍCH ÚLOH .....	8
2.1.1 GEOMETRIE .....	8
2.1.2 MĚŘÍCÍ PROFILY .....	9
2.1.3 TYPY PROUDĚNÍ .....	9
2.1.4 DIMENZE .....	9
2.1.5 ASYNCHRONICITA .....	10
2.2 VOLBA MODELU .....	10
2.3 PARAMETRIZACE ZÁKLADNÍCH ÚLOH A SIMULACÍ .....	11
2.3.1 INTENZITA GENEROVÁNÍ AGENTŮ .....	11
2.3.2 ŠÍŘKA KORIDORU .....	11
2.3.3 PARAMETRY AGENTŮ .....	11
2.3.4 KALIBRACE MODELU .....	13
2.4 SIMULACE A POSTPROCESSING DAT .....	13
3 KOMFORTU POHYBU OSOB .....	16
3.1 SOUČASNÉ POZNÁNÍ .....	16
3.2 DEFINICE KOMFORTU OSOB .....	17
3.3 VÝPOČET KOMFORTU NA SIMULAČNÍCH DATECH, EXPORT PRO DATABÁZI .....	20
4 DATABÁZE PRO STROJOVÉ UČENÍ .....	22
5 METODY STROJOVÉHO UČENÍ A JEJICH APLIKACE .....	24
5.1 PREPROCESSING .....	26
5.1.1 KÓDOVÁNÍ NOMINÁLNÍCH HODNOT .....	26
5.1.2 STANDARDIZACE DAT .....	26
5.1.3 PŘÍKLAD KÓDOVÁNÍ DAT .....	26
5.2 VÝBĚR ALGORITMU .....	27
5.2.1 UMĚLÉ NEURONOVÉ SÍTĚ .....	27
5.3 NÁVRH ARCHITEKTURY – HYPERPARAMETER TUNING .....	29

5.4	Metriky vyhodnocení.....	31
6	VÝVOJ SOFTWAREVÉHO NÁSTROJE.....	32
6.1	HLAVNÍ OKNO.....	32
6.2	SCHÉMA ÚLOHY.....	33
6.3	PARAMETRY ÚLOHY.....	34
7	ZÁVĚR.....	35
	PŘÍLOHY .....	36
	PŘÍLOHA A – METRIKY VYHODNOCENÍ .....	36
	REFERENCE.....	38

## SEZNAM POUŽITÝCH ZKRATEK

---

3D	Trojrozměrný, kartézská soustava souřadnic
AI	(= artificial intelligence), umělá inteligence
ANN	Artificial Neural Network
BIM	Building information model
CAD	Computer-Aided Design
ČR	Česká republika
ČSÚ	Český statistický úřad
I_uni/L_uni/ T_uni_LL	Označení základní úlohy
ID	(= identification), identifikace ve výpočetní technice
MDPP	Model dopravy pěšího proudu
ML	Machine Learning (strojové učení)
MSE	Odmocnina střední kvadratické chyby
P1	Označení profilu MODELU
R2	Koeficient determinace
RMSE	Odmocnina střední kvadratické chyby
RMSLE	Odmocnina střední kvadratické logaritmické chyby

## TERMÍNY A DEFINICE

---

agent	entita reprezentující osobu v simulaci/modelu
clustering	shluková analýza
hmota jádra	integrál přes (hustotní) jádro (přes nějakou oblast)
hustotní jádro	individuální hustotní distribuce = funkce definující rozložení chodce v prostoru
komfort	veličina popisující míru pohodlí (viz metodika)
kongesce	místo s vysokou dynamickou hustotou
osoba	jedinec reprezentující realitu v základní úloze, vizi apod.
Pathfinder	softwarový nástroj pro modelování pěší dopravy
prostorové rozmazání	parametr popisující velikost chodce v prostoru
PySimpleGUI	modul určený pro tvorbu grafického rozhraní v jazyce Python
Python	programovací jazyk
síťka geometrie	diskretizace prostoru
trénovací dataset	dataset užívaný pro natrénování algoritmu umělé inteligence

## 1 ÚVOD

Již při sestavování prvotního stavebního návrhu si tvůrci kladou za cíl zajistit maximální komfort při užívání navrhované stavby jejich budoucími uživateli. V praxi se nicméně ukazuje, že ne vždy se tohoto záměru v projekční fázi povede docílit. Nebo naopak se setkáváme s případy, že samotní správci a provozovatelé svými dodatečnými, neuváženými změnami a rozmisťováním zařízení/poutačů/ mobiliáře apod. nerespektují původní záměr a nastavení komfortu pro běžné fungování konkrétní budovy.

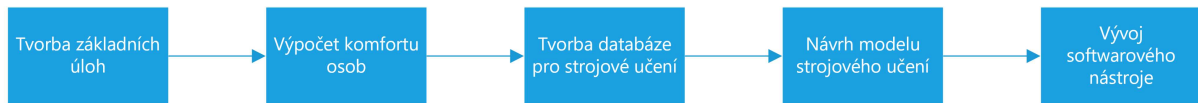
Komfort je definován jako veličina, která na základě metody hustotních jader stanovuje statistický odhad lokální hustoty v rámci předepsané doby. Pro tuto veličinu je v následujícím textu uveden podrobný postup výpočtu a jeho predikce na základě časoprostorových trajektorií agentů, získaných z běžných modelovacích nástrojů. Predikce slouží jako ukazatel, který je zaměřený na běžný provoz budov.

Základní opatření je možné aplikovat již v projekční fázi, a to úpravou základní geometrie návrhu budovy – zejména dimenzemi koridorů, zúžených míst a rozptylových ploch, což napomáhá k přirozenému navýšení průtoku chodců při běžném provozu či každodenních aktivitách a tím i zvýšení jejich komfortu a ochotě setrvat v těchto prostorách. Na základě vyhodnocení komfortu je možné pomoci při návrhu a dimenzování konkrétních vnitřních dispozic navrhovaných budov, díky čemuž je možné docílit i jeho zvýšení při pohybu osob v jednotlivých typech prostorů a zvětšit i schopnost flexibilně predikovat vliv případných změn/úprav při provozní reorganizaci interiéru.

S ohledem na to byla vytvořena i tato metodika, která se zabývá časoprostorovými predikcemi komfortu na umělých datech z numerických modelů pohybu osob. Tato metodika je určena všem výzkumným týmům, osobám zabývající se predikcemi a inženýrům z praxe, kteří chtějí pro svou práci používat právě časoprostorové predikce, jejich testování a validaci tak, jak tomu bylo i v našem projektu *TJ04000232 - Efektivní časoprostorové predikce s využitím metod strojového učení* primárně pro určování komfortu, ale i z globálního hlediska pro jiná zaměření.

Uživatel by měl mít základní povědomí o numerických modelech pohybu osob, základní znalosti o strojovém učení a práci s databázemi. Metodika seznamuje uživatele s jednotlivými fázemi procesu časoprostorových predikcí komfortu osob a jejich dílčími procesy, jako jsou:

- tvorba základních úloh v návaznosti na vlastní řešené prostředí,
- volba modelu a jeho parametrizací,
- simulace a post-processing dat,
- definice a výpočet komfortu osob,
- sestavení vlastní databáze,
- aplikace metod strojového učení (příprava dat, výběr vhodného algoritmu, návrh architektury algoritmu a jeho metrik vyhodnocení),
- vývoj softwarového nástroje.



Obr. 1 - Workflow metodiky časoprostorových predikcí komfortu osob (zdroj: vlastní).

Tento postup nabízí efektivní řešení, jak vytvořit funkční predikční model na bázi neuronových sítí pro sadu tzv. „základních úloh“ a eliminovat tak čas a náklady spojené s nutností tvorby simulací numerického modelu v mnoha variacích v reálném čase při ověřování stavby (ať už v přípravné projekční fázi či provozních úpravách).

## 2 ZÁKLADNÍ SROVNÁVACÍ ÚLOHY A JEJICH PARAMETRIZACE

Sestavení základní sady úloh geometrie vychází ze samotného návrhu posuzované stavby, který by měl již reflektovat základní potřeby komfortního pohybu osob vycházející především z předpisů a norem zaměřující na typologii staveb stanovující dimenze pro návrh staveb na daném území. V České republice se jedná zejména o následující dokumenty:

- **NEUFERT, Ernst. Navrhování staveb. Praha: CONSULTINVEST, 1995. 581 s. [1],**
- **Vyhláška č. 398/2009 Sb. o obecných technických požadavcích zabezpečujících bezbariérové užívání staveb [2],**
- **Vychází z PBR — ČSN 73 0802 (ČSN 73 08XX) [2].**

Historicky i s přesahem do dnešní doby je drtivá většina staveb navrhována v pravoúhlém uspořádání interiérů – místností, koridorů, funkčních celků apod. S ohledem na tento fakt i nepřímo vyplývá právě samotná volba základních úloh – I, L, T, X. Sada úloh je vytvářena a sestavována s ohledem na přirozený pohyb osob v prostoru. Jeho principy lze dělit na čtyři charakteristické podskupiny typů základních úloh:

- **koridor** (pro ověření dimenzí a návrhu zejména z pohledu šířkových nároků a manipulačních ploch - např. i s ohledem na osoby s hendikepem),
- **křížení** (pro ověření návrhu jeho provozu a vlivu na pohyb osob při křížení proudů),
- **překážky** (pro ověření vlivu a vzniku dynamických změn v davu),
- **schodiště** (pro ověření dimenzí a návrhu zejména z pohledu šířkových nároků a jejich uspořádání, s ohledem na to, že z provozního hlediska se jedná o prvky geometrie s nejnáročnějším popisem pohybu osob).

Samotná řešená stavba je pro zjednodušení dělena na základní dílčí celky, které nám následně představují jednotlivé základní úlohy pro analýzu dílčích míst geometrie stavby. Úlohy jsou voleny na základě potřeby ověření dimenzí, návaznosti a míst, kde může docházet ke kongescím, dynamickým zvrátům pohybu osob či jejich střetávání a mísení. Příklad výběru základních úloh je specifikován v Tab. 1. Jedná se o základní úlohy typu koridor a křížení. Tyto úlohy byly dále děleny do podskupin, které jsou v tabulce specifikovány. Výběr úloh pokrývá základní dispoziční uspořádání budov v rámci jednoho podlaží, je však možné ho libovolně rozšířit o další výše zmíněné typy.

Tab. 1 - Příklad výběru a dílčího členění základních úloh

Název	Popis úlohy
I_uni	úloha typu I s jednosměrným prouděním osob
I_bidi	úloha typu I s obousměrným prouděním osob
I_bidi_asyn	úloha typu I s asynchronním obousměrným prouděním osob
L_uni	úloha typu L s jednosměrným prouděním osob
L_bidi	úloha typu L s obousměrným prouděním osob
L_bidi_asyn	úloha typu L s asynchronním obousměrným prouděním osob

<b>T_IL_uni</b>	úloha typu T s jednosměrným prouděním osob, kde pouze jeden proud zatáčí o 90°
<b>T_IL_bidi</b>	úloha typu T s obousměrným prouděním osob, kde pouze jeden proud zatáčí o 90°
<b>T_IL_bidi_asyn</b>	úloha typu T s asynchronním obousměrným prouděním osob, kde pouze jeden proud zatáčí o 90°
<b>T_LL_uni</b>	úloha typu T s jednosměrným prouděním osob, kde oba proudy zatáčí o 90°
<b>T_LL_bidi</b>	úloha typu T s obousměrným prouděním osob, kde oba proudy zatáčí o 90°
<b>T_LL_bidi_asyn</b>	úloha typu T s asynchronním obousměrným prouděním osob, kde oba proudy zatáčí o 90°
<b>X_uni</b>	úloha typu X s jednosměrným prouděním osob
<b>X_bidi</b>	úloha typu X s obousměrným prouděním osob
<b>X_bidi_asyn</b>	úloha typu X s asynchronním obousměrným prouděním osob

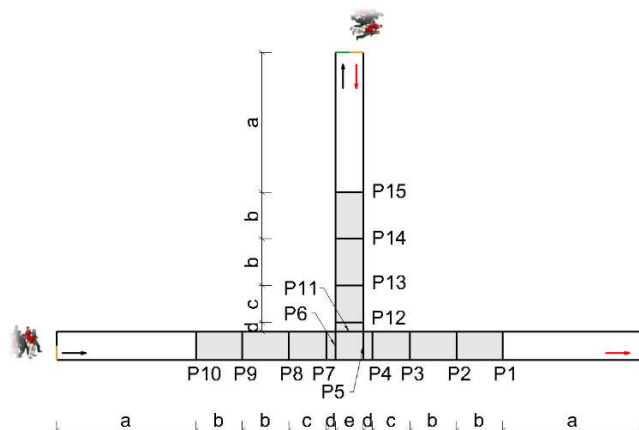
Základní úkoly jsou následně modelovány v numerickém modelu pohybu osob. V našem případě byl pro modelování užít nástroj SW Pathfinder (verze 2022.1.0422 x64) [4], viz podkap. 2.2 výběr modelu. Z výsledků dílčích modelů jsou filtrována a zpracovávána data, která jsou následně využívána jako podklad pro výpočty a statistiky nově zavedené veličiny tzv. “komfortu” (blíže o definici komfortu a jeho výpočtu v kap. 3.3). Z vypočtených a analyzovaných dat je dále sestavována databáze, která následně slouží jako trénovací (ověřovací) množina pro implementaci vybrané metody ML, jak je blíže uvedeno v kap. 4.

## 2.1 TVORBA ZÁKLADNÍCH ÚLOH

V případě, že zvolený model dopravy pěšího proudu (dále MDPP) umožňuje import modelu geometrie ve formě BIM, případně CAD souborů s příponou .dwg, .dxf, .ifc, .dae nebo .fbx, je doporučeno geometrii základních úloh předpřipravit v programech, které jsou oproti grafickým rozhraním MDPP určeny pro rychlé a přesné technické kreslení (např. AutoCAD, Revit, apod.). Při tvorbě geometrie i následném importu do MDPP je vhodné pro efektivní práci zachovat stejné jednotky a počátek v souřadném systému napříč užitými nástroji.

### 2.1.1 GEOMETRIE

Geometrii základní úlohy je třeba sestavit tak, aby oblast měření komfortu byla ustálená (= pozorovaný úsek) a aby v ní docházelo pouze k jevům dynamiky pohybu osob, které lze reálně očekávat. Například uvažujeme-li úlohu s obousměrným prouděním, viz Obr. 2, kde je vstupní kalibrační profil napojovaného úseku úlohy T rozdělen na vstup a výstup (černá a červená šipka) a generátor agentů by navazoval bezprostředně na měřící profil P15, byly by do výpočtu komfortu započítány kolize, ke kterým může dojít při generování agentů a vzájemnému vyhýbání při vstupu a výstupu obou proudů. Tyto kolize nesouvisí s běžným pohybem osob v prostředí pozorovaného úseku základní úlohy a není tedy žádoucí, aby ovlivňovaly výpočet komfortu. Pro předejití těmto situacím doporučujeme před pozorovaný úsek (šedá barva) vložit koridory (bílá barva), ve kterých dojde k ustálení proudů osob před vstupem do pozorované oblasti.



Tabulka rozměrů [m]	
a	15.00
b	5.00
c	4.00
d	1.00
e	dle zadané šířky koridoru

Obr. 2 - Schéma příkladu geometrie základní úlohy typu T (vlevo) před vstupem do MDPP s rozmístěním a označením měřících profilů, pozorovaného úseku (šedá barva) s obousměrným prouděním a tabulka s dimenzemi (vpravo).

## 2.1.2 MĚŘÍCÍ PROFILY

V geometrii základních úloh je zapotřebí definovat konkrétní souřadnice měřících profilů, kde bude komfort osob později počítán. Do samotné geometrie není potřeba profily zadávat, jelikož by mohly mít negativní vliv na trajektorie pohybu jednotlivých agentů. Souřadnice profilů vstupují do procesu až při výpočtu samotného komfortu. Rozmístění a počet profilů by měl takový, aby byl obsažen celý pozorovaný úsek.

V místech, kde dochází ke zvýšenému výskytu kongescí či změnám směrů geometrie – např. u úloh typu L, T, X a v oblastech rohů geometrie je doporučeno zvýšit hustotu rozmístění měřících profilů s cílem zachytit veškeré jevy spjaté s dynamickými změnami a vznikem kongescí. Příklad geometrie základní úlohy s rozmístěním měřících profilů v koridorech je znázorněn na Obr. 2.

## 2.1.3 TYPY PROUDĚNÍ

Základní úlohy je potřeba rozlišit nejen z hlediska geometrie, ale i typu proudění, které může nastat – jednosměrné nebo obousměrné. Při jednosměrném proudění je pohyb osob obecně jednodušší a lze při něm dosáhnout výrazně vyšších intenzit než při pohybu obousměrném, kde dochází častěji ke kolizím. Samotný algoritmus pro řešení kolizí je závislý na technologii konkrétně zvoleného modelu a může mít vliv na celkové výsledky, proto je zapotřebí, aby při tvorbě základních úloh byly zohledněny specifika vybraného modelu (blíže v kap. 2.2 Výběr modelu).

## 2.1.4 DIMENZE

Dimenze základních úloh jsou převážně variovány z hlediska šířky koridoru. Úlohy by měly zahrnovat standardně užívané rozměry s rovnoměrným krokem (např. 0,5 m). Čím jemnější krok je volen, tím přesnější lze očekávat i následné predikce v rámci strojového učení. Nelze však, ale očekávat lineární nárůst přesnosti, spíše konvergenci k maximálně možné dosažitelné hodnotě. Zároveň s jemnějším krokem enormně roste objem trénovacích dat.

## 2.1.5 ASYNCHRONICITA

Při obousměrném proudění může nastat situace, při které se intenzity obou proudů liší – proudy jsou asynchronní. Základní úlohy s obousměrným typem proudění jsou členěny z hlediska asynchronicity. Asynchronní úlohy jsou samostatným typem základních úloh. Míru asynchronicity lze stanovit např. procentuálním stupňováním nejvyšší intenzity pro daný typ úlohy, jak je uvedeno v Tab. 2.

Tab. 2 - Příklad definování asynchronicity vycházející z úlohy s obousměrným prouděním o maximální intenzitě generování agentů  $0,9 \text{ os.s}^{-1}$ .

Asynchronicita [%]	Intenzita generování v 1. proudu [ $\text{os.s}^{-1}$ ]	Intenzita generování v 2. proudu [ $\text{os.s}^{-1}$ ]
15	$0.9 - (0.9 \times 0.15) = 0.77$	$0.9 + (0.9 \times 0.15) = 1.04$
30	$0.9 - (0.9 \times 0.30) = 0.63$	$0.9 + (0.9 \times 0.30) = 1.17$
45	$0.9 - (0.9 \times 0.45) = 0.50$	$0.9 + (0.9 \times 0.45) = 1.31$
60	$0.9 - (0.9 \times 0.60) = 0.36$	$0.9 + (0.9 \times 0.60) = 1.44$

## 2.2 VOLBA MODELU

Z hlediska komplexnosti jsou základní úlohy zřejmé a zahrnují jednosměrný či obousměrný pohyb osob v rámci koridoru. Zde použitý MDPP Pathfinder je průběžně validován na experimentálních datech, jejichž výsledky jsou doložitelné [5]. V rámci validačních testů byl pozorován především vztah mezi hustotou a rychlostí osob. Z analýzy vyplynulo, že pohybový algoritmus modelu relativně přesně replikuje reálná pozorování.

Pro stanovení komfortu pohybu osob v rámci základních úloh je nutné věnovat pozornost především vhodnému výběru modelu dopravního proudu pěších. Při volbě modelu je nutné dbát na možnosti ztvárnění geometrie i samotné možnosti výpočtu. Model by měl splňovat několik základních podmínek uvedených níže:

1. **Mikroskopické ztvárnění geometrie** – v modelech umožňující toto ztvárnění jsou osoby uvažovány jako jednotlivci (agenti) s vlastní sadou parametrů. Výstupem simulací jsou zejména trajektorie těchto agentů v předepsaných časových krocích simulace, které jsou pro výpočet komfortu klíčové.
2. **Stochasticita** – chování osob a jejich pohyb je ze své podstaty stochastický a stále máme velké množství proměnných parametrů pro jejich popis, které nejsou zcela "vyčísleny", proto je i vhodné přirozenou proměnlivost procesu zohlednit. Trajektorie osob jsou vždy unikátními díky vlivu rozhodovacího procesu osob, chování a interakcí mezi jednotlivci a jejich individuální parametrizací apod. Základní stochasticita je v MDPP vnášena pravděpodobnostním rozdělením jednotlivých parametrů osob, které jsou ověřeny experimentálními měřeními (zejména rychlost pohybu, fyzické rozměry, mobilita apod.). To, že je stochasticita modelu obsažena, poznáme tak, že při opakované realizaci simulace modelu se stejnými parametry nikdy nedosáhneme naprosto stejných výsledků, jelikož při každé nové simulaci základní úlohy dojde k novému přerozdělení parametrů

jednotlivým agentům vycházejícím z počátečně zadaných distribucí. Sledované výsledky (např. komfort nebo doba průchodu) tedy zahrnují směrodatnou odchylku  $> 0$ .

3. **Parametrizace rychlosti pohybu** – jedná se o základní parametr MDPP. Standardně bývá rychlost pohybu zadávána jako normální rozdělení.
4. **Parametrizace fyzických rozměrů osob** – fyzické rozměry osob (zejména šířka ramen) ovlivňují tvorbu kongescí. Distribuce šířky ramen v populaci odpovídá normálnímu rozdělení a pro její zadání lze vyjít z publikace [12][13].
5. **Parametrizace odstupů od stěn a ostatních agentů** – tyto parametry určují, jak velké vzájemné rozestupy a rozestupy od stěn mají agenti dodržovat. V tuto chvíli není k dispozici dostatek experimentálních dat pro přesné stanovení distribuce těchto parametrů a je možné tedy přistoupit k jejich odhadu.

## 2.3 PARAMETRIZACE ZÁKLADNÍCH ÚLOH A SIMULACÍ

Mezi základní parametry MDPP pro analýzu komfortu pohybu osob se řadí parametry jako: **intenzita generování agentů, šířka koridoru úlohy a parametry agentů.**

### 2.3.1 INTENZITA GENEROVÁNÍ AGENTŮ

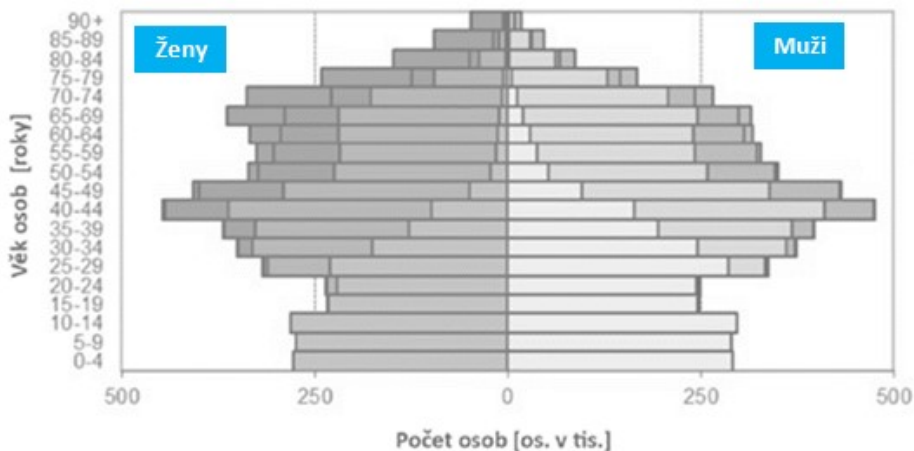
Tento parametr zohledňuje počet vygenerovaných agentů za sekundu z „generátoru agentů“, který je umístěn mimo geometrii základní úlohy, viz Obr. 2 a Obr. 4. Intenzity by měly variovat od nízkých hodnot (při nichž dochází k minimálnímu počtu interakcí) k maximálním hodnotám, které hraničí s tvorbou kongescí v rámci celého pozorovaného úseku. Zde je potřeba zdůraznit, že tvorba kongescí může být do určité míry ovlivněna konkrétním algoritmem pohybu osob (např. řešením kolizí), a je proto potřeba maximální intenzitu kalibrovat pro zvolený model pohybu osob a nepřebírat automaticky hodnoty použité v jiném modelu. Obecně by intenzity měly být zastoupeny alespoň v 5 variantách s nastaveným krokem, přičemž tento krok se může lišit v rámci jednotlivých úloh.

### 2.3.2 ŠÍŘKA KORIDORU

Šířka koridoru by měla zahrnovat minimální a maximální možnou projektovou šířku koridoru z hlediska norem ČSN a mezilehlé šířky s předepsaným krokem. Na základě rozdílu mezi maximální a minimální šířkou je potřeba zvolit krok  $i$  s ohledem na generovaný objem dat. V případě volby minimální hodnoty 1,5 m a maximální hodnoty 3,0 m, které zohledňují většinu dimenzí koridorů užívaných v pozemním stavitelství, se jako optimální krok může zvolit hodnota 0,5 m (celkem tedy 4 varianty).

### 2.3.3 PARAMETRY AGENTŮ

Je vhodné uvažovat parametry odpovídající lokálním podmínkám daného objektu (např. v objektu typu škola bude velký podíl dětí a mladistvých). Sběr lokálních dat je vhodné realizovat analýzou videozáznamu metodami strojového učení (např. analýza rychlostí pohybu na základě trasování pohybu osob, analýza odstupů atd.). V případě, že tyto technologie nejsou dostupné, je možné provést místní šetření či experiment.



Obr. 3 Genderové rozdělení mužů a žen na území ČR, (zdroj: vlastní, přepracováno z [6])

V rámci parametrizace modelu je možné uvažovat distribuci genderu v podobě 49,1 % mužů a 50,9 %, která odpovídá datům z ČR z roku 2019 dle ČSÚ [6]. Mezi další významné parametry patří vzájemné minimální **odstupové vzdálenosti mezi agenty a minimální odstupové vzdálenosti mezi agenty a geometrií**. Pro oba tyto parametry existuje velmi málo empirických dat a jsou v modelu zohledněny konstantou. Vybraný model Pathfinder zahrnuje řadu dalších parametrů souvisejících například s řešením kolizí, tzn. střetu dvou agentů a řešení jejich vzájemného vyhnutí, a další, které nesouvisí s podstatou projektu. Algoritmus řešení kolizí může ovlivnit výpočet komfortu a může se pro různé modely lišit. Je tedy potřeba k parametrům modelu přistupovat individuálně, dle technické reference daného vývojáře. Významné parametry agentů z hlediska projektu jsou popsány v Tab. 3.

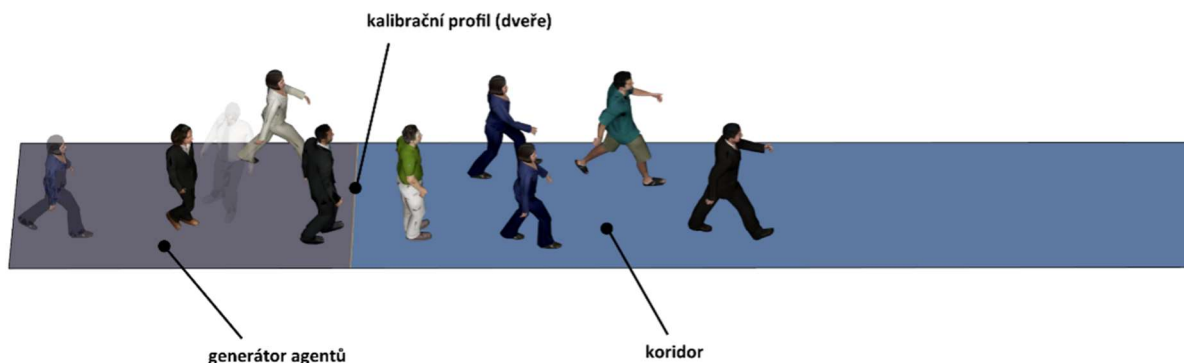
Tab. 3 - Příklad volby parametrů agentů

Parametr	Gender	Konst.	Rovnoměrné rozdělení		Normální rozdělení			
			min	max	$\mu$	$\sigma$	min	max
Šířka ramen	Muži	x	x	x	45,47	2,03	42,12	48,81
	Ženy	x	x	x	39,12	2,03	35,78	42,46
Rychlost pohybu	Muži	x	1,410	1,514	x	x	x	x
	Ženy	x	1,255	1,371	x	x	x	x
Odstupy agentů	Muži	0,08	x	x	x	x	x	x
	Ženy	0,08	x	x	x	x	x	x
Odstupy agentů a geometrie	Muži	0,15	x	x	x	x	x	x
	Ženy	0,15	x	x	x	x	x	x

### 2.3.4 KALIBRACE MODELU

V mikroskopickém modelování je kalibrace procesem, při němž dochází ke změně parametrů modelu s cílem dosažení maximální shody s reálným očekáváním.

Intenzita proudění agentů na profilu uvnitř koridoru je parametr, který lze spočítat inženýrským výpočtem na základě hustoty a rychlosti pohybu, např. dle [5]. Patří mezi základní parametry, na jejichž základě je predikován komfort pohybu. V rámci modelu základních úloh byl kalibrován generátor agentů, viz Obr. 4. Cílem je nastavit optimální intenzitu generování agentů  $[os.s^{-1}]$ , pro kterou je na kalibračním profilu naměřena požadovaná intenzita ve stejných jednotkách. Při nízkých hodnotách intenzit platí, že hodnota generování se rovná intenzitě na profilu. Při vyšších intenzitách se však hodnoty mohou lišit a je potřeba dále kalibrovat. V rámci projektu se kalibrace týkala základních úloh typu "I\_uni" a "L\_uni".



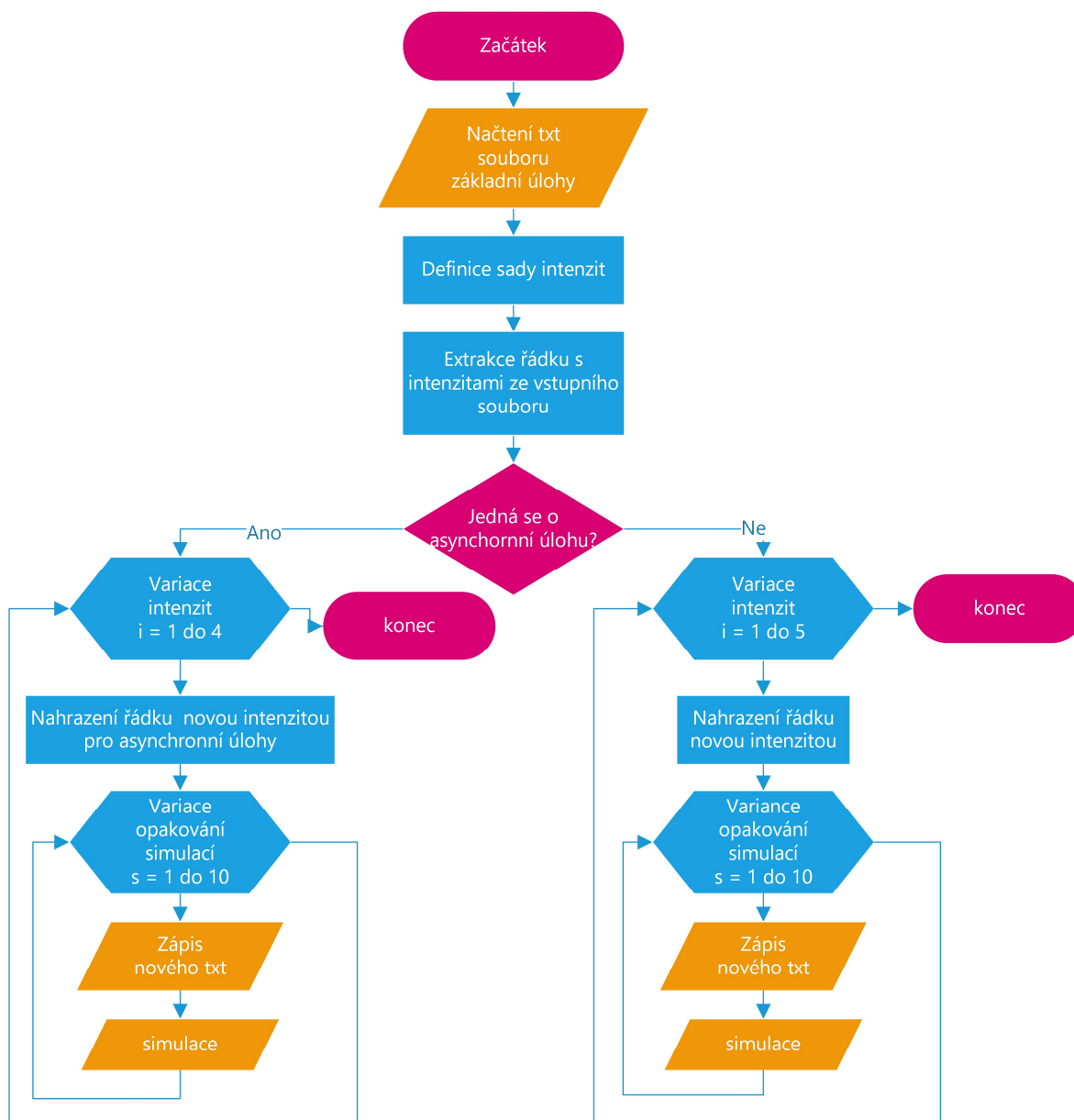
Obr. 4 - Příklad napojení generátoru agentů na koridor a kalibrační profil

## 2.4 SIMULACE A POSTPROCESSING DAT

Simulace modelu základních úloh je vhodné algoritmicky automatizovat, umožňuje-li to vybraný MDPP. Úroveň algoritmizace závisí na schopnostech uživatele a pro samotné testování a validaci není podmínkou nezbytně nutnou. Pathfinder, tedy MDPP nasazený v rámci projektu, je složen z grafického rozhraní, simulačního jádra a 3D prohlížeče výstupů. V grafickém rozhraní probíhá tvorba geometrie a parametrizace jednotlivých variant modelu základních úloh a veškeré údaje o modelu jsou následně uloženy do souboru ve formátu txt. Tento soubor je následně možné automaticky upravit algoritmem před zavedením do simulačního jádra. Mezi doporučené algoritmické úpravy patří automatické generování sady vstupních txt souborů s variantní změnou:

- Typu úlohy,
- Intenzity generování agentů,
- Pořadovým číslem simulace.

Každou variantu nastavení simulace je potřeba simulovat alespoň v 10 opakováních pro zohlednění stochasticity modelu (viz kap. 2.2).



Obr. 5 - Příklad algoritmu pro automatizaci simulací (Zdroj: vlastní).

Na Obr. 5 je znázorněn příklad algoritmu automatizace simulací s variabilním typem úlohy, intenzitou generování agentů a pořadovým číslem simulace. Na začátku dojde k načtení textového souboru, do něž je zapsán model ve fázi, kdy je vytvořena geometrie modelu a zadána parametrizace. Následně je definován vektor s intenzitami generování agentů, které mají být simulovány. Ze vstupního souboru je extrahována řádka, kde je intenzita generování zapsána. Následuje podmínka, která určuje, zda se jedná o asynchronní typ úlohy, kde jsou 4 variace intenzit pro asynchronicitu 15, 30, 45 a 60 %. Pokud se nejedná o asynchronní úlohu, intenzita má 5 variací. Pro každou variaci intenzit se následně generuje 10 upravených textových souborů, které se následně přes příkazovou řádku zavádí do simulátoru.

V případě Pathfinderu jsou výstupní soubory simulací s trajektoriemi agentů ukládány v souborech ve formátu .csv. Před samotným výpočtem je vhodné provést filtraci dat tak, aby byla zachována data pro časy odpovídající ustálenému stavu.

### 3 KOMFORTU POHYBU OSOB

#### 3.1 SOUČASNÉ POZNÁNÍ

Výzkum v oboru aplikací pěších proudů, zejména evakuací, zpravidla užívá Fruinem zavedenou metriku *úrovně kvality pěší dopravy*, která klasifikuje lokální hustotu [os.m<sup>-2</sup>] na stupnici A-F. Z hlediska fyzikálních veličin je vhodné komfort pohybu chodců analyzovat z pohledu okamžité lokální hustoty, ale i z hlediska jejího vývoje v čase. Zároveň je důležité, aby byl do výpočtu zahrnut i psychologický aspekt (např. v podobě proxemických zón, viz kap. 3.2), který standardní výpočet hustoty ze své podstaty zohlednit nemůže.

Pro definici komfortu je využita lokální hustota definovaná pomocí hustotních jader. Každý chodec je uvažován jako zdroj *jádra (individuální hustotní distribuce)*, která může mít libovolný vhodný tvar (např. kužel, válec nebo Gaussovo distribuci) [9][10][11]. Každé takové jádro  $p_i(\vec{x})$  pro chodce  $i$  má parametr  $R$ , který popisuje *rozměr (prostorové rozmazání) chodce*. Zmíněná jádra jsou volena normovaná na jedničku, proto každý chodec může do celkové hustotní distribuce přispět maximálně jedničkou, tj. libovolným reálným číslem z intervalu [0, 1]. *Lokální hustotu v oblasti A* lze poté zapsat s využitím standardní definice [9][10][11] jako:

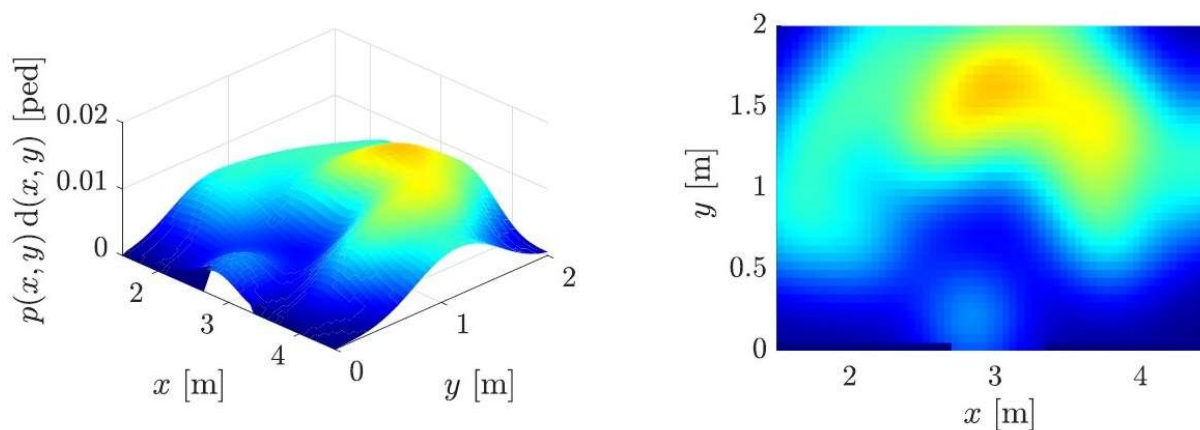
$$\rho_A = \frac{N}{|A|} = \frac{\int_A p(\vec{x}) d\vec{x}}{|A|} = \frac{\int_A \sum_{i=1}^N p_i(\vec{x}) d\vec{x}}{|A|}, \quad (1)$$

kde  $p(\vec{x}) = \sum_{i=1}^N p_i(\vec{x})$  je hustotní distribuce v celé oblasti (tj. součet všech jader příslušejících chodcům v pozorované oblasti). Příklad hustotní distribuce je uveden na Obr. 6. Pokud oblast  $A$  označuje celý pozorovaný prostor, pak:

$$\int_A \sum_{i=1}^N p_i(\vec{x}) d\vec{x} = N, \quad (2)$$

kde  $N$  značí celkový počet chodců v pozorované oblasti (tj. v oblasti  $A$  jsou všichni chodci).

Ačkoliv je pohyb chodců spojitý, při výpočtu komfortu je potřeba použít časoprostorovou diskretizaci, tj. do zkoumané oblasti v rovině  $x$ - $y$  je zavedena síťka se čtverečky o délce strany  $xStep = 0,05$  m a čas postupuje v krocích o délce  $tStep = 1$  s.



Obr. 6 - Hustotní distribuce s Gaussovým jádrem – vlevo: pohled z boku, vpravo: pohled shora (Zdroj: vlastní).

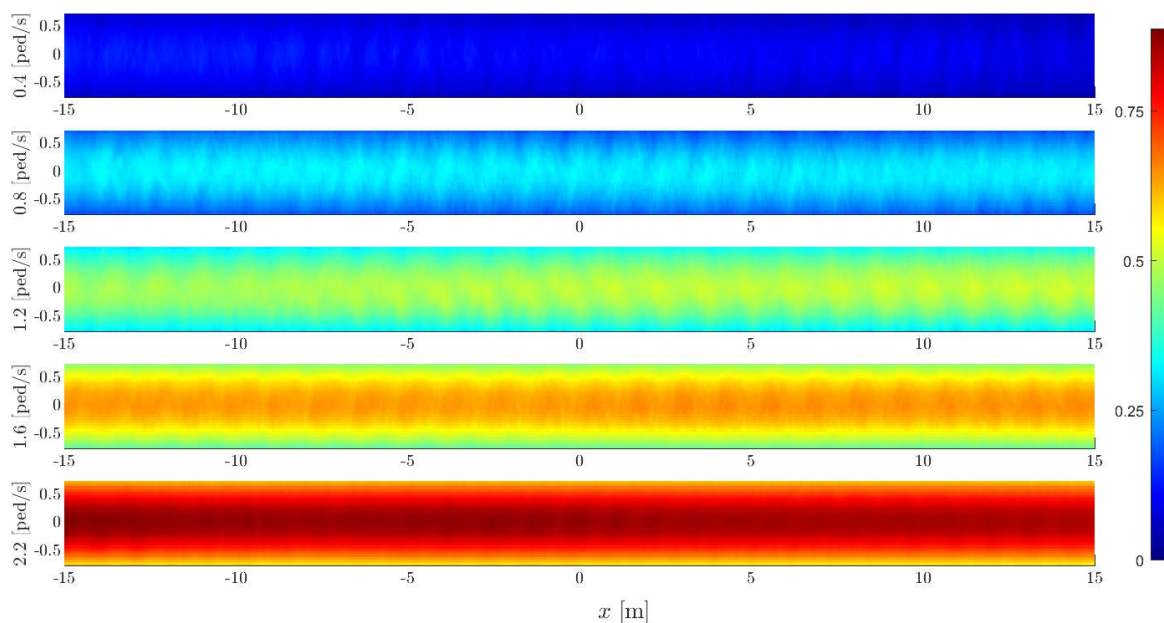
### 3.2 DEFINICE KOMFORTU OSOB

Pro definici komfortu je voleno Gaussovo jádro následovně:

$$p_i(\vec{x}) = \frac{1}{\sqrt{2\pi R^2}} \exp\left(-\frac{\|\vec{x}-\vec{x}_i\|^2}{2R^2}\right), \quad (3)$$

kde  $\vec{x}_i$  značí pozici chodce  $i$ . Toto jádro je pro tyto účely nevhodnější díky svým matematickým (hladká a klesající funkce) i socio-fyzikálním vlastnostem (demonstruje vliv chodce na ostatní v neomezeném okolí, tj. chodci se vzájemně ovlivňují i na velké vzdálenosti, i když jen minimálně). V každém čase a v každém bodě roviny  $x$ - $y$  je posléze vyhodnocena lokální hustota  $\rho_A(\vec{x}, t)$  dle vztahu (1), kde oblast  $A$  je jeden čtvereček sítky geometrie. Komfort pro každý bod této roviny je poté definován jakožto medián přes časovou proměnnou, tj.

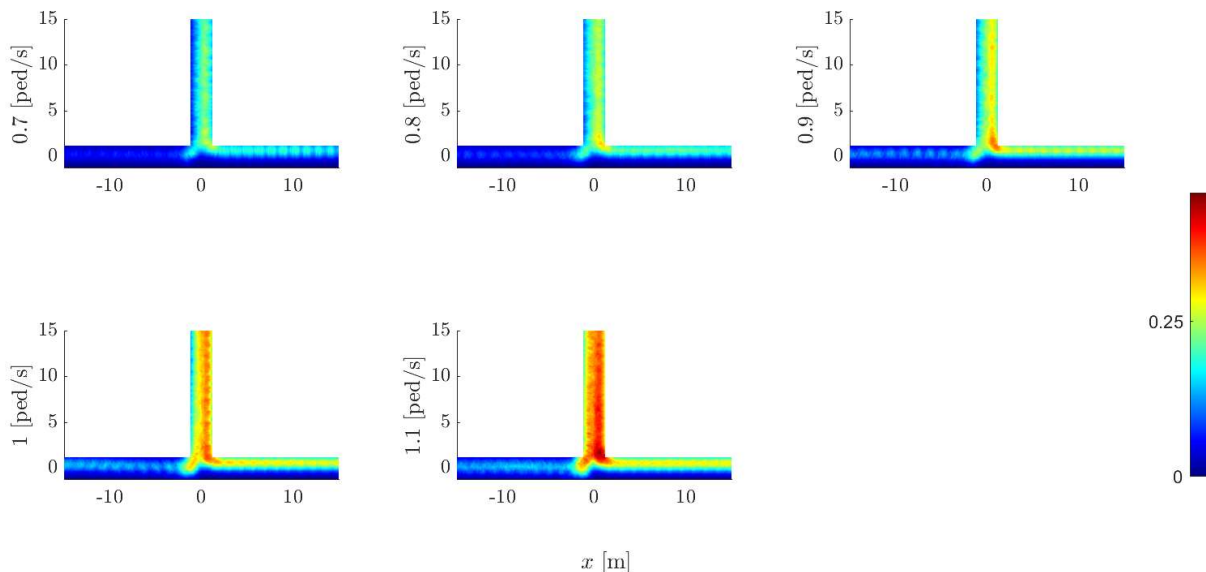
$$\text{komfort}(\vec{x}, R) = \rho_{0.5}(\vec{x}, R) = \text{Median}_{t \in R^+} \rho(\vec{x}, R, t), \quad [\text{os} \cdot \text{m}^{-2}]. \quad (4)$$



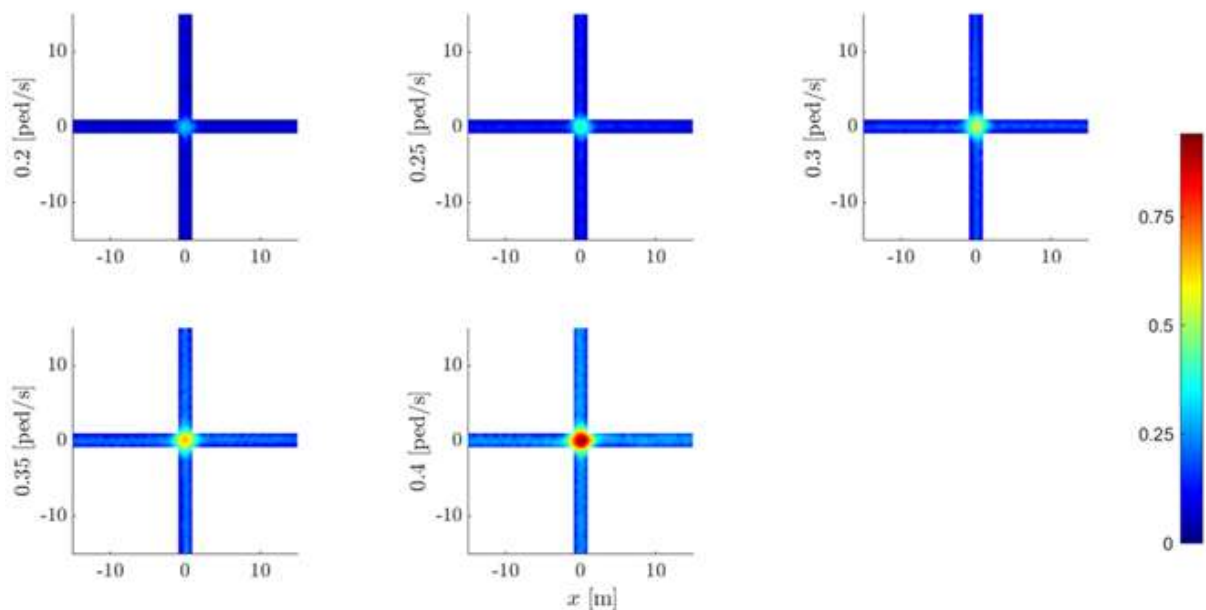
Obr. 7 - Příklad veličiny komfort  $(x, R)$   $[\text{os}/\text{m}^2]$  pro základní úlohu typu I s jedním proudem a šířkou koridoru 1500 mm pro různé vstupní intenzity. (Zdroj: vlastní.)

Zpracovali  
 Ing. Martina Kratochvílová  
 Ing. Petra Okřínová  
 Ing. Ondřej Uhlík  
 Ing. Jana Vacková  
 Ing. Jiří Apeltauer, Ph.D.

Datum 30. 05. 2022

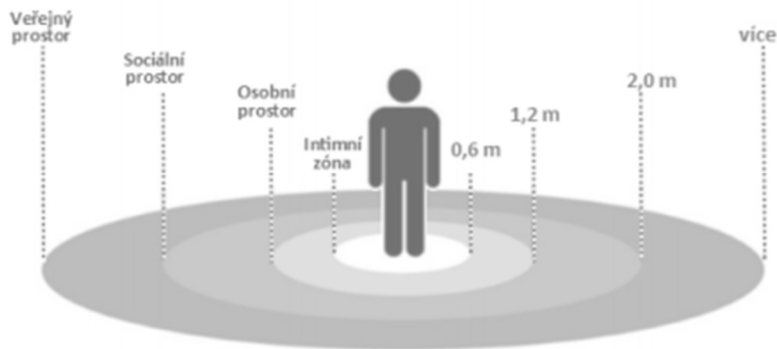


Obr. 8 - Příklad veličiny komfort( $x, R$ ) [ $os/m^2$ ] pro základní úlohu typu T se dvěma proudy (LL) a šířkou koridoru 2500 mm pro různé vstupní intenzity. (Zdroj: vlastní.)



Obr. 9 - Příklad veličiny komfort ( $x, R$ ) [ $os/m^2$ ] pro základní úlohu typu X se dvěma proudy a šířkou koridoru 2000 mm pro různé vstupní intenzity. (Zdroj: vlastní.)

Volba velikosti Gaussova jádra vychází z teorie proxemiky [11][12], která pramení ze 4 základních zón založených na vzdálenostech mezi lidmi: intimní, osobní, společenské a veřejné (znázorněno na Obr. 10). Velikost jádra je zde nastavena tak, aby 80 % jeho hmoty spadalo do osobní a intimní zóny a zbytek (20 %) jádra do ostatních zón.



Obr. 10 - Proxemické zóny osoby užitá pro diagram dle Edwarda T. Halla z r. 1966. [12][13][14]

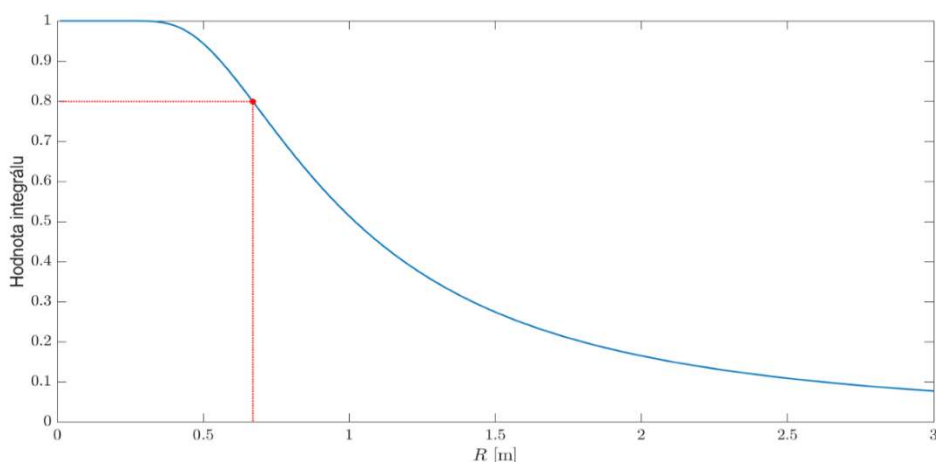
Toho se dosáhne vyřešením následující rovnice:

$$\int_{A_{R(PS)}} p(\vec{x}, R) d\vec{x} = 0,8, \quad (5)$$

kde oblast

$$A_{R(PS)} = \{\vec{x} \in \mathbf{R}^2: |\vec{x}| \leq R(PS)\} \quad (6)$$

definuje plochu intimní a osobní zóny z Obr. 3 a funkce  $p(x, R)$  představuje Gaussovo jádro, které je pro zjednodušení lokalizované do bodu  $[0,0]$ . Optimální hodnota pak činí pro  $R(PS) = 1,2$  m (nastavenou podle Obr. 4) hodnota  $R = 0,67$  m, kdy dochází k tomu, že 80 % hmoty jádra je uvnitř kruhu o poloměru  $R(PS)$ . Křivka znázorňující hodnotu integrálu je k vidění na obrázku Obr. 5 (z grafu lze též vyčíst hodnotu  $R$  pro případnou hodnotu jinou než zde nastavených 80 %). Příklady takto vypočtených komfortů pro různé základní úlohy je k vidění na Obr. 7, Obr. 8 a Obr. 9.



Obr. 11 - Vyhodnocení integrálu (Zdroj: vlastní).

### 3.3 VÝPOČET KOMFORTU NA SIMULAČNÍCH DATECH, EXPORT PRO DATABÁZI

#### Input pro výpočet komfortu

- Trajektorie z osob z numerického modelu (např. Pathfinder).
- Všechny zkoumané parametrické sady (zn. *paramSet*).

Fields	type	stream	substream	width	intensity	asyn
1	'I'	'uni'	NaN	1500	0.4000	NaN
2	'I'	'uni'	NaN	1500	0.8000	NaN
3	'I'	'uni'	NaN	1500	1.2000	NaN
4	'I'	'uni'	NaN	1500	1.6000	NaN
5	'I'	'uni'	NaN	1500	2.2000	NaN

Obr. 12 - Příklad tabulky *paramSet* definující všechny zkoumané parametrické sady.

- Počet iterací každé parametrické sady uvedené v *paramSet*.
- Délka prostorového kroku pro síťku položenou do roviny x-y (zn. *xStep*).
- Délka časového kroku (zn. *tStep*).
- Parametry *minT* a *maxT* definující ustálený stav systému v intervalu  $[minT, maxT]$ .
- Síťka s vytyčenou geometrií místnosti (např. koridor, křižovatka) dle *paramSet*.
- Počáteční a koncový bod všech zkoumaných profilů (zn. *profileSetup*).

Fields	type	width	profile_name	first_point	second_point
1	'I'	1500	'P1'	[-10,0.7500]	[-10,-0.7500]
2	'I'	1500	'P2'	[-5,0.7500]	[-5,-0.7500]
3	'I'	1500	'P3'	[0,0.7500]	[0,-0.7500]
4	'I'	1500	'P4'	[5,0.7500]	[5,-0.7500]
5	'I'	1500	'P5'	[10,0.7500]	[10,-0.7500]
6	'I'	2000	'P1'	[-10,1]	[-10,-1]
7	'I'	2000	'P2'	[-5,1]	[-5,-1]

Obr. 13 - Příklad tabulky *profileSetup* definující pozice profilů pro všechny zkoumané geometrie.

#### Postup výpočtu komfortu

1. Nastavení parametrů: *jádro*, velikost jádra *R*.
2. Výpočet komfortu pro zkoumanou geometrii (její každý bod z místnosti) pro všechny parametrické sady uvedené v *paramSet*:

##### Pro každou iteraci:

- Preprocessing dat z Pathfinderu:
  - Tvorba souboru se sloupečky: *ID chodce*, souřadnice *x*, souřadnice *y*, čas *t*, tj. na každém řádku je jeden chodec a jemu příslušné vektory (vše oříznuté dle *minT* a *maxT*) - zn. *pedPath*.

Fields	ID	x	y	t
1	27	[21.7959,23.1492,24.5024,25]	[0.4770,0.4770,0.4770,0.4770]	[100,101,102,103]
2	28	[16.8262,18.1821,19.5380,20....]	[-0.2296,-0.2296,-0.2296,-0.22...]	[100,101,102,103,....]
3	29	[15.8883,17.2442,18.6001,19....]	[0.3729,0.3729,0.3729,0.3729,0...]	[100,101,102,103,....]
4	30	1x13 double	1x13 double	1x13 double
5	31	1x15 double	1x15 double	1x15 double

Obr. 14 - Příklad tabulky *pedPath* vytvořené na základě dat (trajektorii) z *Pathfinderu*.

- Tvorba souboru s časovými řezy, tj. se sloupečky čas  $t$  (v rozmezí  $minT$  a  $maxT$ ),  $ID$  chodců v daném čase  $t$  v místnosti, souřadnice  $x$  všech chodců v místnosti v daném čase  $t$ , souřadnice  $y$  všech chodců v místnosti v daném čase  $t$  – zn. *pedTime*.

Fields	t	ID	x	y
1	100	15x1 double	15x1 double	15x1 double
2	101	15x1 double	15x1 double	15x1 double
3	102	16x1 double	16x1 double	16x1 double

Obr. 15 - Příklad tabulky *pedTime* vytvořené na základě dat (trajektorii)u z *Pathfinderu*.

- Pro každý čas  $t$ :
  - Do všech pozic  $[x_i, y_i]$  aktuálních chodců v místnosti v čase  $t$  dle *pedTime* je umístěno jádro.
  - Výpočet hustoty v každém bodě místnosti  $[x, y]$  dle (1).
- Výpočet komfortu dle vztahu (2).

Výsledkem tohoto kroku je datový soubor *comfort2D* se sloupečky  $ID$  parametrické sady dle *paramSet*,  $ID$  iterace, *comfort* (pro každý bod v dané geometrii).

- Napočítání statistik dle volby uživatele, tj. ze souboru *comfort2D* jsou pro všechny parametrické sady dle *paramSet* s využitím všech iterací vypočteny dílčí statistiky, vhodné jsou např. průměr, medián, horní/dolní kvantil, rozptyl apod.
- Výpočet komfortu pro zkoumané profily:
  - Dle *profileSetup* vyfiltrujeme z výsledků bodu 2, resp. 3 pomocí sítky geometrie danou profilovou úsečku a hodnoty v jejích bodech uložíme do souboru *comfortProfile*.
  - Výstupem je soubor *comfortProfile* ve formátu .csv, který je pak vstupem do databáze/neuronové sítě.

## 4 DATABÁZE PRO STROJOVÉ UČENÍ

Výsledné hodnoty komfortu pro zvolené základní úlohy, profily a kombinace parametrů je vhodné ukládat do strukturované podoby – databáze (strukturovaná podoba je klíčová pro další strojové zpracování výsledků), která může mít například formu .csv tabulky nebo databázové aplikace. Databáze mohou být lokální, např. ve formě Microsoft Access, nebo vzdálené (SQL, MySQL, atd.). Volba konkrétní podoby vychází z volby základních úloh a záleží na celé řadě faktorů, jako je počet kombinací, počet uživatelů, kteří budou s výsledky pracovat, jejich geografická poloha atd. Všechny varianty však sdílí společný základní prvek, kterým je vektor výsledných hodnot komfortu na daném profilu pro danou kombinaci prvků. Tento vektor má dvě části, popisnou a hodnotovou. Popisná část musí jednoznačně identifikovat, o jakou kombinaci vstupních parametrů jde a o jaký vektor se jedná. Hodnotami popisné části tak mohou být např.:

1. type – typ úlohy (I, L, T, ...),
2. stream – typ proudění (uni, bidi, ...),
3. substream – podtyp proudění pro úlohu K (IL, LL, ...),
4. width – šířka koridoru [mm],
5. intensity – intenzita pěšího proudu [ $\text{os}\cdot\text{s}^{-1}$ ],
6. profile\_name – unikátní identifikátor profilu.

Následně hodnotová část vektoru obsahuje samotné hodnoty v jednotlivých bodech profilu:

7. mean\_profile\_comfort\_1 – průměrná hodnota komfortu na daném profilu pro bod 1,
8. mean\_profile\_comfort\_2 – průměrná hodnota komfortu na daném profilu pro bod 2,
9. ...

U hodnotové části, respektive při definování základních úloh, je nutné stanovit obecné globální pravidlo pro orientaci profilů a jejich číslování tak, aby bylo zřejmé, kde daný bod na daném profilu leží (jestli je např. číslován zleva doprava nebo naopak – Obr. 2). Alternativou může být doplnění popisné části o informaci, ze které půjde orientaci profilu odvodit (např. pravidlo pravé ruky po proudu osob).

Jednotlivé vektory je vhodné ukládat do tabulek a ty následně do databází nebo na disk ve formátu .csv, .xlsx atd. Tabulka pak může vypadat například následovně:

Tab. 4 - Ukázka struktury databáze, část - T\_bidi\_LL\_stats.csv

type	stream	substream	width	intensity	profile_name	mean_profile_comfort_1	median_profile_comfort_1	IQR_profile_comfort_1	Q3_profile_comfort_1	Q1_profile_comfort_1	var_profile_comfort_1	midrange_profile_comfort_1	CV_profile_comfort_1	robustCV_profile_comfort_1
T	bidi	LL	1500	0,20	P1	0,008	0,008	0,001	0,009	0,007	0,000	0,007	0,139	0,088
T	bidi	LL	1500	0,20	P2	0,005	0,005	0,001	0,005	0,004	0,000	0,005	0,165	0,133
T	bidi	LL	1500	0,20	P3	0,005	0,005	0,001	0,005	0,004	0,000	0,005	0,123	0,085
T	bidi	LL	1500	0,20	P4	0,005	0,006	0,001	0,006	0,005	0,000	0,005	0,142	0,083
T	bidi	LL	1500	0,20	P5	0,012	0,012	0,002	0,013	0,011	0,000	0,012	0,115	0,092
T	bidi	LL	1500	0,20	P6	0,013	0,013	0,002	0,014	0,012	0,000	0,014	0,086	0,062
T	bidi	LL	1500	0,20	P7	0,004	0,003	0,001	0,004	0,003	0,000	0,005	0,238	0,132
T	bidi	LL	1500	0,20	P8	0,003	0,003	0,001	0,003	0,002	0,000	0,003	0,228	0,178
T	bidi	LL	1500	0,20	P9	0,003	0,003	0,001	0,004	0,003	0,000	0,003	0,155	0,106
T	bidi	LL	1500	0,20	P10	0,002	0,002	0,001	0,003	0,002	0,000	0,002	0,337	0,273
T	bidi	LL	1500	0,20	P11	0,119	0,116	0,022	0,133	0,110	0,000	0,121	0,095	0,091
T	bidi	LL	1500	0,20	P12	0,100	0,100	0,007	0,105	0,097	0,000	0,097	0,075	0,036
T	bidi	LL	1500	0,20	P13	0,075	0,075	0,009	0,080	0,072	0,000	0,074	0,080	0,056
T	bidi	LL	1500	0,20	P14	0,075	0,075	0,011	0,081	0,070	0,000	0,074	0,097	0,071
T	bidi	LL	1500	0,20	P15	0,075	0,074	0,006	0,078	0,072	0,000	0,075	0,068	0,040

## 5 METODY STROJOVÉHO UČENÍ A JEJICH APLIKACE

Algoritmy ML se primárně dělí podle formy trénovací množiny na učení s učitelem a bez učitele, jejich kombinaci, a v poslední době často používané zpětnovazebné učení. Při učení s učitelem stroji pomáhají lidé a trénovacím datům na vstupu je přiřazen správný výstup. Pokud však řešitel (člověk) není schopen určit správnou hodnotu výstupu, používá se učení bez učitele. Zpětnovazebné učení je založeno na principech behaviorální psychologie a je často používáno u agentních modelů. V ML je nejčastěji používáno metodou Markovova rozhodovacího procesu a jejími variantami.

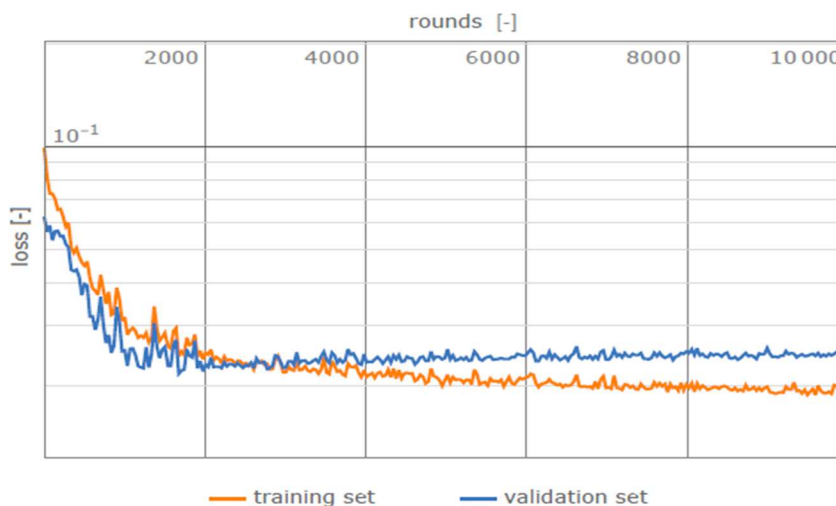
ML je obvykle užíváno na rozdělování vstupních dat do tříd (klasifikace), určení hodnoty na výstupu dle hodnoty na vstupu (regrese), případně k seskupování objektů do skupin s podobnými vlastnostmi (clustering). Dále se různé metody ML používají ke zpracování řeči nebo textu, vyhodnocování obrázků, a další [19][20].

Jak již bylo zmíněno (a je i patrné z typu základních úloh), ML vychází převážně ze statistiky. V současnosti je používáno mnoho různých metod, jejich kombinací a další se neustále vyvíjejí.

V případě učení s učitelem obvyklý postup pro užití metod strojového učení sestává z následujících kroků:

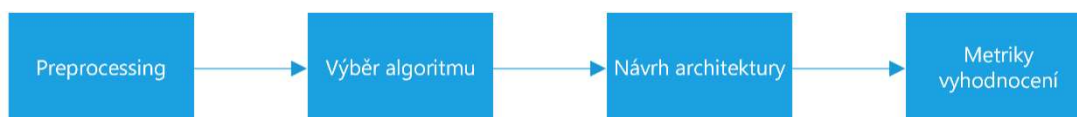
- **Určení prediktorů** – V současnosti, takřka ve všech oblastech, produkujeme mnohem větší množství dat, než jsme schopni zpracovávat. Je tedy běžnou situací, že ne všechny proměnné, které dostaneme, jsou nutné pro řešení našeho problému. Mnohdy jejich užití může být dokonce kontraproduktivní – čím více vstupujících proměnných, tím vyšší dimenzionalita nebo šance zmatení modelu. Vhodné je tedy provést například citlivostní analýzu.
- **Transformace dat do vhodného formátu a měřítka** – strojové učení může být aplikováno na různé typy dat (numerické, časové řady, obrázky, ...). Před jejich zpracováním modelem je však nutné, tato data převést na vhodný formát. Tím je obvykle vektor numerických hodnot. Nejčastěji je také vyžadováno, aby jednotlivé vektory byly stejné délky. Stejně tak bývá obvykle vhodné provést standardizaci vstupujících proměnných do podobného měřítka, tak aby některé z proměnných nebyly například 100krát větší než jiné.
- **Volba typu úlohy** – jedná se o regresní nebo klasifikační problém? Ne vždy je odpověď na tuto otázku jednoznačná. Někdy může být výhodnější použití regresního modelu na problém rozdělení dat do dvou tříd (tedy 0 a 1). Výstupem takového modelu potom není jasně daná odpověď 0 nebo 1 (děj nastane/nenastane), ale číslo mezi 0 a 1. Takový výsledek pak může být chápán jako pravděpodobnost, se kterou děj nastane. To může být výhodné, pokud například je pro nás větším rizikem jedna nebo druhá možnost – můžeme nastavit prahovou hodnotu.
- **Volba algoritmu strojového učení** – obvykle prováděno empiricky srovnáním několika algoritmů a výběrem toho nejvýhodnějšího. Nutno dopředu stanovit kritéria výběru s ohledem nejen na přesnost, ale například na výpočetní a časovou náročnost.

- **Volba architektury modelu** – u většiny algoritmů je nutné zvolit hyperparametry, které tvoří jakousi architekturu model. Tyto parametry se během učení nemění. K automatizaci hledání jejich hodnot slouží optimalizační metody (Hyperparametr Tuning).
- **Sestavení trénovacích a testovacích podmnožin** – běžný poměr dělení dat je 80 (70): 20 (30) ve smyslu trénovací: testovací. Data v obou podmnožinách by měly odpovídat rozdělení dat. Proto je doporučováno před rozdělením provést zamíchání dat. Pro většinou algoritmů je vhodné, když trénovací datasety pro jednotlivé třídy jsou přibližně stejné velikosti (stejný počet dat pro každou třídu v setu).
- **Učení** – při učení jsou automaticky upravovány parametry modelu tak, aby co nejlépe popisoval data. Jsou však i algoritmy, které nemají fázi předchozího učení (metoda k-nejbližších sousedů). Proces učení je ovlivněn volbou některých hyperparametrů. V případě neuronových sítí je to například počet epoch (kolikrát je pro učení použit celý trénovací dataset) a velikost dávky (po jak velkém množství dat jsou aktualizovány hodnoty parametrů).
- **Testování** – model je nutné validovat a ověřit si nejen jeho výkon na datech, na kterých se učil, ale i jeho schopnost problém zobecnit. To je prováděno na testovacích datech. Dobrým ukazatelem a srovnáním, jak si model vede na známých a nových datech během učení, jsou křivky učení. Ukazují hodnotu námi zvolené metriky (např. Root Mean Squared Error – RMSE) pro trénovací i testovací datasety v jednotlivých epochách. Metrika by měla být volena dle typu úlohy a závažnosti chyby prvního a druhého druhu.



Obr. 16 - Příklad křivky učení. Z grafu je možné vyčíst, že model je nedostatečně natrénován [Zdroj: vlastní].

Diagram na Obr. 17 popisuje proces tvorby modelu strojového učení. Proces je následně specifikován v následujících podkapitolách.



Obr. 17: Proces tvorby modelu strojového učení

## 5.1 PREPROCESSING

### 5.1.1 KÓDOVÁNÍ NOMINÁLNÍCH HODNOT

Většina modelů strojového učení vyžaduje, aby vstupující a vystupující proměnné byly numerické. Pokud jsou tyto proměnné nominální, tedy i kategorické (například třída, typ úlohy), je nutné je zakódovat tak, aby byly pro model "srozumitelné". Nejběžněji používané metody kódování se dělí na ordinální (řadové) a One-Hot.

#### Ordinální kódování

Při tomto kódování je každé třídě přiřazeno unikátní celé číslo. Běžně jsou používána čísla začínající nulou. Pokud kategorická data mají smysluplné pořadí, může být toto kódování vhodné. Problém však může nastat, pokud mezi třídami není žádný ordinální vztah (např. barva nebo jiné vlastnosti) a my ho tam tímto způsobem zavedeme. Pokud bychom jednotlivým třídám přiřadili pouze po sobě jdoucí celá čísla, například třída 4 by byla pro model 4násobkem třídy 1, atd. V tomto případě může být takové kódování nedostatečné nebo, ještě hůř, zmatečné pro náš model (např. predikce na pomezí tříd). To je i případ základních úloh, jejichž kategorie mezi sebou nemají žádný ordinální vztah.

#### One-Hot kódování

V tomto případě jsou třídy nahrazeny binární proměnnou (např. modrá [1,0,0], červená [0,1,0], zelená [0,0,1]). Je však zřejmé, že takto zakódovaná data obsahují redundanci (např. [0,0,0]). Tento problém řeší Dummy encoding, který  $C$  tříd nahradí  $C-1$  binárními hodnotami. V dříve zmiňovaném případě by třídy byly: [1,0], [0,1], [0,0]. Výhodou tohoto přístupu je také snížení dimenzionality kódované proměnné.

### 5.1.2 STANDARDIZACE DAT

Vstupní proměnné mohou mít různé jednotky (např. metry, kilometry a hodiny), což dále může znamenat, že proměnné mají různá měřítka. Rozdíly v měřících mezi vstupními proměnnými mohou zvýšit obtížnost modelovat daný problém. Příkladem může být, kdy velké vstupní hodnoty (např. rozpětí stovek nebo tisíců jednotek) mohou vést k modelu s velkými hodnotami vah. Takový model je často nestabilní, dlouho se učí a je citlivý na vstupní hodnoty, což má za následek vyšší chybu zobecnění. Nepsaným pravidlem je, že hodnoty vstupující do modelu by měly být malé (absolutní hodnota mezi 0 a 1).

1. **Normalizace dat** – změna měřítka každého proměnné na hodnoty od 0 po 1.
2. **Standardizace dat** – změna měřítka distribuční funkce proměnné tak, aby průměr pozorovaných hodnot byl 0 a směrodatná odchylka byla 1.

### 5.1.3 PŘÍKLAD KÓDOVÁNÍ DAT

V následující Tab. 5 je uveden příklad kódování základních úloh. Na prvním řádku jsou vstupy z databáze, druhý zobrazuje kódování jednotlivých hodnot a třetí vstupní vektor pro model po provedení standardizace.

Tab. 5 - Příklad postupného kódování základních úloh

	Typ úlohy								Šířka	Inten.	Profil	Bod
<b>Vstup</b>	I_bidi_stats								2000	0.2	P3	1
<b>Kódování</b>	1	0	0	0	0	0	0	0	2000	0.2	1	0
<b>Standardizace</b>	1	0	0	0	0	0	0	0	0.3333	0.0185	0.1429	0

## 5.2 VÝBĚR ALGORITMU

Výběr vhodné metody strojového učení je jedním z klíčových kroků ML. Bohužel však neexistuje žádné pravidlo, které by nám říkalo, jaká metoda se pro daný problém hodí nejvíce. Stejně tak zatím nebyla vytvořena žádná metodika pro výběr této metody. Běžnou praxí je, že se stanoví vhodná metrika pro porovnání několika metod (viz. kap. 5.4 Metriky vyhodnocení), provede se vyhodnocení této metriky pro jednotlivé modely se základními (neoptimalizovanými) parametry a následně je vybrána nejvhodnější metoda. Její parametry jsou pak následně optimalizovány. Aby byla co nejlépe popsána výkonnost modelu na nových datech, je obvykle používán některý druh resamplingu (např. k-násobná křížová validace).

**K-násobná křížová validace** je rozdělení dat na  $k$  podmnožin. Testovaný model je pak  $k$ -krát učen tak, že vždy jedna z podmnožin je použita pro testování a zbylé jsou použity pro trénování. Pro každý cyklus učení je stanovena hodnota námi zvolené metriky. Tyto hodnoty jsou obvykle na konci zprůměrovány (nebo je zvolena jiná statistika) a výsledná hodnota (nebo hodnoty) slouží jako popisná charakteristika výkonnosti daného modelu.

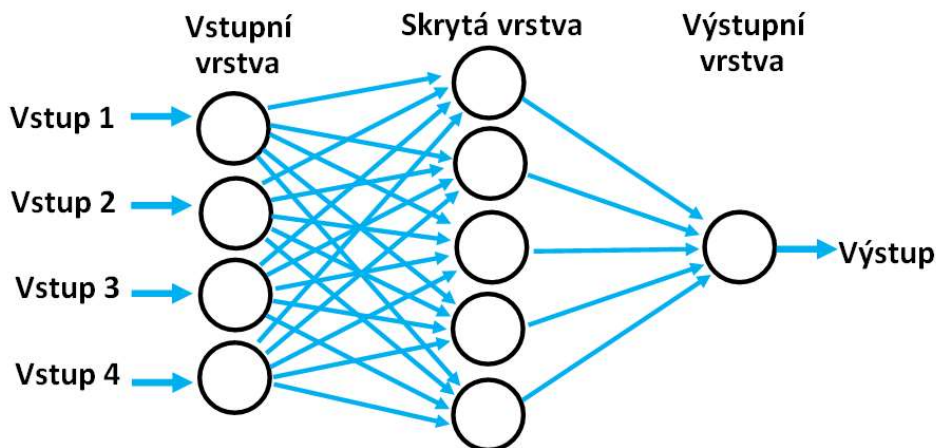
Tab. 6 - Příklad srovnání 4 modelů pomocí metriky RMSE

Metoda	RMSE
Polynomická regrese	0,362678
SVM Gaussovské jádro	0,355443
SVM Polynomické jádro	0,362499
Neuronová síť	0,427711

### 5.2.1 UMĚLÉ NEURONOVÉ SÍTĚ

Nejznámější a také pravděpodobně nejuniverzálnější je metoda neuronových sítí (Artificial Neural Network – ANN). Tento algoritmus je inspirován rozhodovacím procesem lidského mozku. Ten je mimo jiné složen z několika milionů neuronů, které vyhodnocují a mezi sebou si předávají informace. Stejně tak umělá neuronová síť je složená z vrstev (layers), které jsou složené z neuronů. Každá vrstva má daná pravidla, na jejichž základě vyhodnocuje vstupní informace z předchozí vrstvy a předává výstup další vrstvě. Vstup a výstup může být různých formátů. V rámci ANN je však informace předávána pomocí vnitřního systému vah. Počet vrstev a neuronů

v nich je libovolný, jako i typ vrstev. Všechny tyto parametry však ovlivňují spolehlivost ANN. Rozhodně se však nedá říct, že čím více, tím lépe. Tím se dostáváme i k hlavnímu problému ANN – přeurčitosti sítě. Tedy že síť, kromě skutečného trendu nebo pravidla, natrénuje i šum, a tak dává zdánlivě správné výsledky.



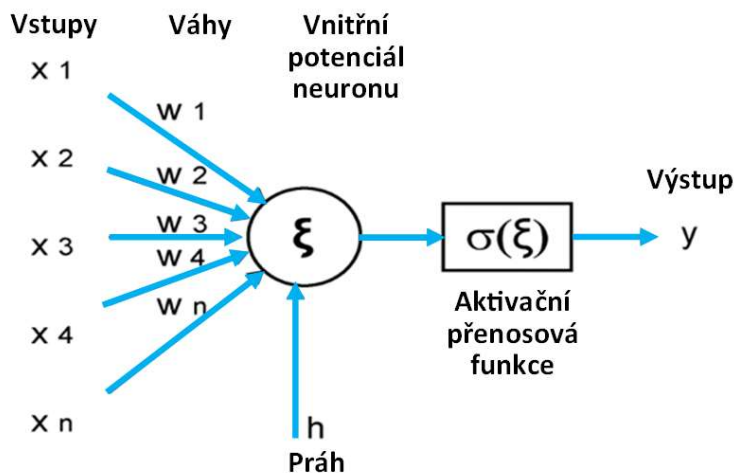
Obr. 18 - Příklad architektury neuronové sítě s jednou skrytou vrstvou [17]

Neuronová síť se skládá z parametrů (naučené) a hyperparametrů (stanovené). Hyperparametry by se daly chápat jako architektura dané sítě a během učení se nemění a o jejich velikosti nebo podobě je nutné rozhodnout na začátku celého procesu. Je to například:

- počet skrytých vrstev a jejich typ,
- počet neuronů v jednotlivých vrstvách,
- Hyperparametry speciálních vrstev (např. drop-out),
- druh optimalizační funkce,
- počet epoch učení,
- typ aktivační funkce,
- velikost dávky (pro každou dávku se aktualizují hodnoty vah),
- ztrátová funkce atd.

Naopak parametry se mění v průběhu učení s cílem zvýšit výkon algoritmu. Jsou to:

1. **Váhy** (= weights) jednotlivých neuronů, které transformují hodnotu do nich vstupující.
2. **Vychýlení** (= bias), které náleží každé vrstvě a síť neustále nepatrně rozvažuje, aby se zvýšila pravděpodobnost nalezení globálního minima ztrátové funkce.



Obr. 19 - Model neuronu [17].

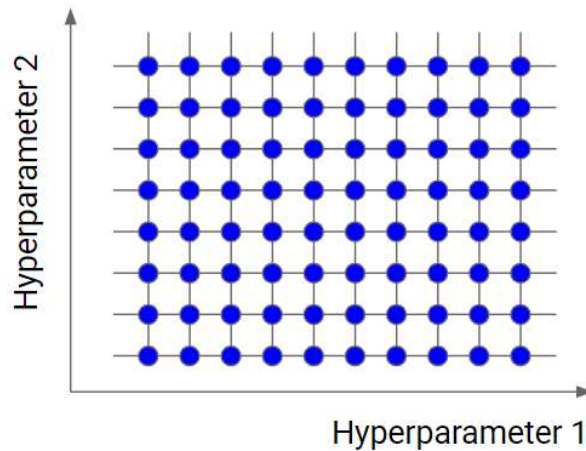
Proces učení probíhá na trénovacích datech, u nichž je znám výstup. Zprvu jsou váhy jednotlivých neuronů iniciovány dle zadané metody a je spočítán výstup a jeho chyba oproti skutečné hodnotě (ztrátová funkce) pro všechna data ve svazku. Svazek může obsahovat od jednoho prvku trénovací množiny až po všechny. Následně je provedena úprava vah pomocí algoritmu zpětného šíření chyby, kdy se, zjednodušeně řečeno, zjišťuje, jak chyba závisí na vahách jednotlivých neuronů. Při zpětném postupu (od výstupu k vstupu) je počítán gradient ztrátové funkce.

### 5.3 NÁVRH ARCHITEKTURY – HYPERPARAMETER TUNING

Hyperparametry jsou hodnoty, které stanovujeme pro model, a které významně ovlivňují jeho výkonnost. Neučí se automaticky během tréninkové fáze, ale musí být poskytnuty explicitně.

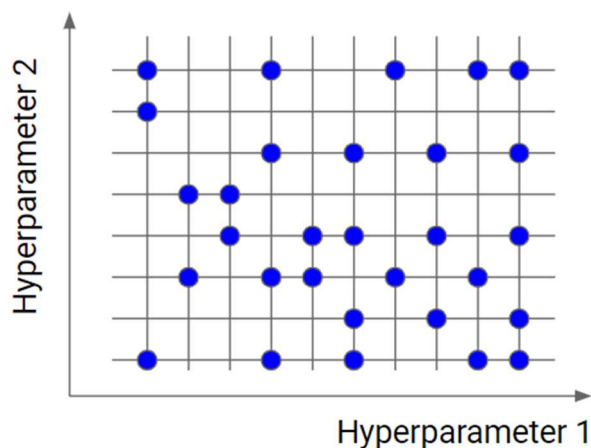
Hyperparametry hrají hlavní roli ve výkonu modelu a měly by být zvoleny a nastaveny tak, aby se přesnost modelu zlepšila. Stanovení je možné provést pomocí dřívější zkušenosti s daným typem úlohy nebo hyperparametry optimalizovat. Není možné používat stále stejnou architekturu pro všechny typy úloh. Optimalizační algoritmy obvykle na hledání hyperparametrů nahlíží jako na minimalizaci nebo maximalizaci nějaké funkce. Pro tento proces je nutné mít sestavenou i validační množinu, tedy data, které model při učení neviděl a je možné na nich ověřovat přesnost. Hyperparametry se vzájemně ovlivňují a je tedy nutné optimalizovat celý jejich set (nelze provádět optimalizaci pro každý hyperparametr zvlášť). Je samozřejmě možné ladění dělat manuálně, nicméně najít tak globální minimum je velmi časově i výpočetně náročné. Globální optimalizace je náročný problém hledání vstupu, jehož výsledkem jsou minimální nebo maximální náklady na danou účelovou funkci. Některé metody pro ladění jsou:

1. **Grid search** (= hledání na mřížce) je nejjednodušší algoritmus pro ladění hyperparametrů. Doména hyperparametrů je rozdělena na diskrétní mřížku. Poté zkusíme každou kombinaci hodnot této mřížky a pomocí křížové validace vypočítáme některé metriky výkonu. Bod mřížky, který maximalizuje (nebo minimalizuje) průměrnou hodnotu křížové validace, je optimální kombinací hodnot pro hyperparametry.



Obr. 20 - Příklad procesu hledání na mřížce [18]

2. **Random Search** (= náhodné vyhledávání) je podobné vyhledávání v mřížce, ale místo použití všech bodů v mřížce testuje pouze náhodně vybranou podmnožinu těchto bodů. Čím menší je tato podmnožina, tím rychlejší, ale méně přesná je optimalizace. Čím větší je tato datová sada, tím přesnější je optimalizace, ale tím blíže k vyhledávání v mřížce.



Obr. 21 - Příklad náhodného vyhledávání [18]

3. **Bayesovská optimalizace** poskytuje techniku založenou na Bayesově teorému

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}. \quad (7)$$

Funguje tak, že sestrojí pravděpodobnostní model účelové funkce, nazývaný náhradní funkce (= Surrogate Function). Tento model nám pak říká, s jakou pravděpodobností je nové řešení lepší než to stávající. V prvním kroku najdeme optimální hyperparametry pro náhradní model. Následně je aplikujeme na skutečnou účelovou funkci. S těmito výsledky aktualizujeme náhradní model. Kroky pak opakujeme, dokud nedosáhneme stanoveného limitu. Základní myšlenkou je: věnujte trochu více času výběru dalších hyperparametrů, abyste mohli provádět méně volání účelové funkce. V praxi je čas

strávený výběrem dalších hyperparametrů bezvýznamný ve srovnání s časem stráveným vyhodnocováním účelové funkce. K výběru nových hyperparametrů slouží akviziční funkce (7).

4. **Hyperband** je optimalizovaná verze náhodného vyhledávání, která využívá včasné zastavení k urychlení procesu ladění hyperparametrů. Hlavní myšlenkou je natrénovat velké množství modelů během malého počtu epoch a pokračovat v učení pouze pro modely dosahující nejvyšší přesnosti na validačním set

## 5.4 METRIKY VYHODNOCENÍ

**Regrese** se týká problémů prediktivního modelování, které zahrnují předpovídání číselné hodnoty. Na rozdíl od klasifikačních problémů není možné použít metriku přesnosti, kolik vzorků bylo správně označeno. Je nutné využít metriky, které různými postupy vyčíslují chybu (vzdálenost) skutečné hodnoty od bodu na regresní přímce. Patří sem například:

1. **Střední kvadratická chyba (MSE)** - střední hodnota druhých mocnin rozdílů mezi skutečnou hodnotou a odhadem. Nevýhodou může být, že jednotky výsledku jsou druhou mocninou, a tedy občas těžko představitelné.
2. **Odmocnina střední kvadratické chyby (RMSE)** - je rozšířením MSE, kdy hodnota MSE je odmocněna. Její výhodou je srozumitelnost jednotek, které jsou v základním tvaru.
3. **Střední absolutní chyba (MAE)** - součet všech absolutních hodnot chyb. Na rozdíl od předchozích metrik nepenalizuje výrazně odlehlé body.
4. **Odmocnina střední kvadratické logaritmické chyby (RMSLE)** - přirozený logaritmus RMSE. Tato metrika je vhodná, pokud očekáváme, že se měřítko chyb bude výrazně lišit. Je tedy mnohem robustnější při vyhodnocování dat s odlehlými hodnotami.
5. **R<sup>2</sup>** - neboli koeficient determinace. Popisuje variabilitu závisle proměnné. Nabývá hodnot mezi 0 a 1. Čím vyšší číslo, tím přesnější model. Velkou nevýhodou je, že nezohledňuje počet proměnných (tzn. neporovnatelné datasey rozdílných velikostí). Tento problém však řeší **adjustovaný koeficient determinace**.

## 6 VÝVOJ SOFTWAREVÉHO NÁSTROJE

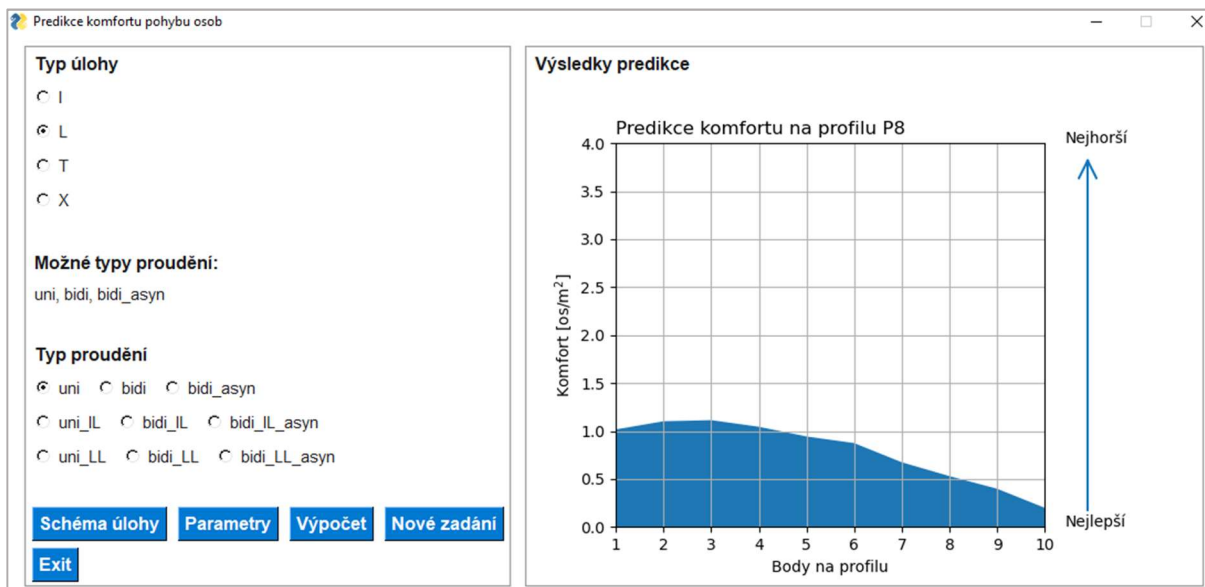
Softwarový nástroj (dále SW) je klíčovým prvkem pro komunikaci mezi výše popsaným systémem predikce komfortu a koncovým uživatelem. Cílem SW je poskytnout okamžitý odhad komfortu dle zadaných parametrů v rámci uživatelsky snadno ovladatelného grafického rozhraní (dále GUI). Vývoj SW lze realizovat v programovacím jazyce Python (verze 3.9), konkrétně v modulu PySimpleGUI. Tento modul je určen právě pro tvorbu GUI. Prvním krokem je stanovení celkové podoby rozhraní, tzn. získání odpovědí na otázky:

- Jak bude prostředí uspořádáno?
- Jakým způsobem budou zadávána vstupní data?
- V jaké formě budou prezentovány výsledky?

V následujících podkapitolách je doporučen postup pro zodpovězení uvedených otázek. Při zadávání vstupních dat by měl mít uživatel k dispozici 2 samostatné komponenty, které může spustit ze základního prostředí (tzv. hlavního okna). V následujících podkapitolách je uveden možný příklad uspořádání komponent pro efektivní a příjemnou práci s daným nástrojem.

### 6.1 HLAVNÍ OKNO

Jedná se o základní komponentu, ve které probíhá komunikace s ostatními komponentami a kde dochází k zobrazení výsledků. Příklad uspořádání hlavního okna je zobrazen na Obr. 22, kde je toto okno rozděleno na dvě části. Pravá část slouží k zobrazení a interpretaci grafu predikce komfortu. V levé části dochází ke specifikaci typu úlohy a typu proudění a otevírání doplňkových oken popsanych níže.



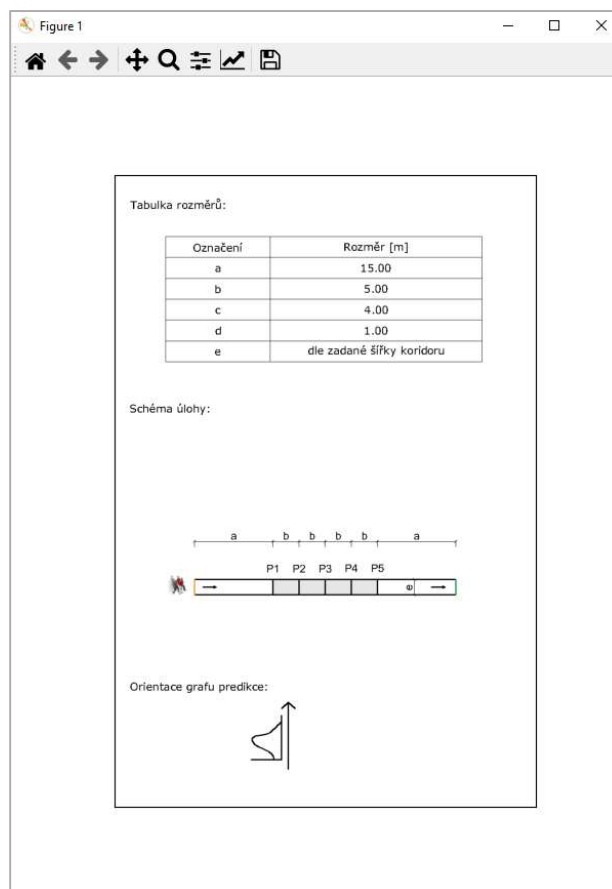
Obr. 22 – Ukázka hlavního okna GUI softwarového nástroje „Komfort“ s příkladem predikce.

Na základě zvoleného typu úlohy se uživateli automaticky zobrazí možnosti volby typu proudění. Obě volby jsou řešeny formou přepínačů. Dále se v levé části hlavního okna nachází celkem 5 tlačítek popsanych níže:

- *Schéma úlohy.* Po stisknutí se uživateli zobrazí vyskakovací okno se schématem zvolené úlohy (viz 6.2).
- *Parametry.* Po stisknutí se uživateli zobrazí vyskakovací okno s parametry základní úlohy (viz 6.3).
- *Predikce.* Po stisknutí dojde k poslání zadané vstupní matice do ANN a zobrazení plošného grafu komfortu v pravé části hlavního okna.
- *Nové zadání.* V případě potřeby další predikce je zapotřebí stisknout toto tlačítko, čímž dojde k vynulování schématu základní úlohy grafu predikce a veškerých zadaných parametrů. Následně může uživatel zadávat nové parametry.
- *Exit.* V případě stisknutí tlačítka dojde k zavření hlavního okna a celého software.

## 6.2 SCHÉMA ÚLOHY

Po stisknutí tlačítka „Schéma úlohy“ dojde k zobrazení okna se schématem úlohy, viz Obr. 23. Komponenta v horní části umožňuje základní práci s obrázkem – jeho posun, zoom, uložení apod. Obrázek obsahuje celkem tři informace: tabulku rozměrů úlohy, okótované schéma s rozmístěním profilů a znázorněným typem proudění a orientaci grafu predikce ve vztahu ke schématu. Šedá oblast značí pozorovaný úsek, v němž jsou umístěny měřící profily, na kterých je komfort počítán. Informace o orientaci bodů na profilech je klíčová pro správné čtení a interpretaci grafu komfortu. Komponentu je možné kdykoli zavřít případně vůbec neotevírat v průběhu celého procesu.

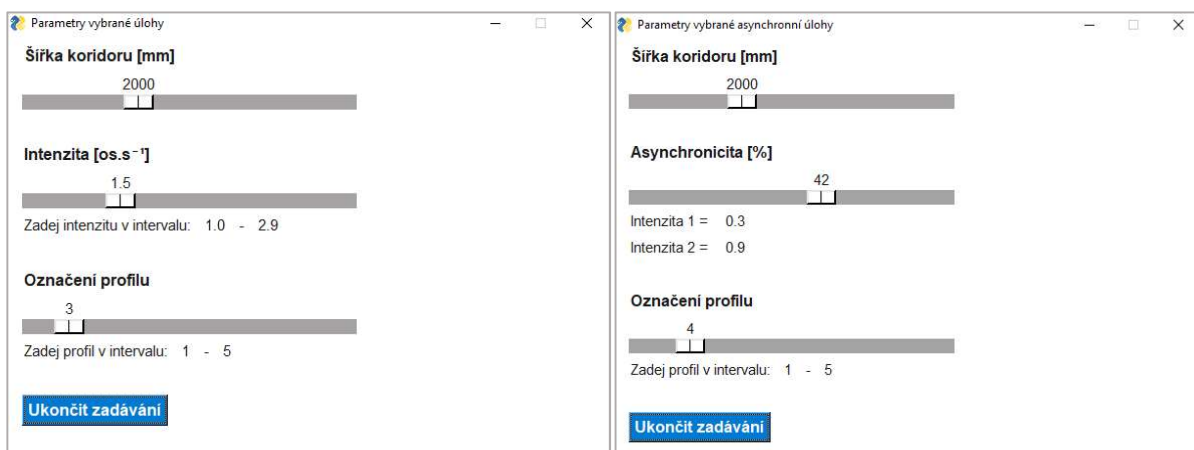


Obr. 23: Komponenta se schématem úlohy.

### 6.3 PARAMETRY ÚLOHY

Zadávání parametrů je možné řešit samostatným oknem, ve kterém dochází k zadání parametrů pro vybraný typ úlohy a proudění. Okna je možné rozlišit podle typu proudění (synchronní nebo asynchronní) vzhledem ke skutečnosti, že u obou typů proudění dochází k zadávání jiných dat. V případě synchronních úloh by měla komponenta obsahovat parametry: šířka koridoru, intenzita a profil, viz Obr. 24. Šířku koridoru je možné zadat v rozsahu 1500–3000 mm. Krok je vhodné nastavit dle zásad projektování pozemních staveb. V rámci výstupu projektu *TJ04000232-V2 Softwarový nástroj pro efektivní časoprostorové predikce založené na ML* byl krok zadání nastaven na 10 cm. Intenzitu je možné zadat v intervalu, který je automaticky počítán lineární interpolací na základě zadané šířky koridoru. Krok zadání intenzity je doporučeno nastavit na 0,1 [os.s<sup>-1</sup>]. Poslední parametr je výběr profilu, kde má být komfort predikován. Komponenta automaticky zobrazuje možný rozsah výběru profilu, na základě vybraného typu úlohy. Je vhodné mít souběžně s touto komponentou otevřenou komponentu se schématem úlohy pro snadnější výběr požadovaného profilu.

Zadávání je vhodné řešit formou posuvníků, čímž je umožněna interpolace hodnot šířky koridoru, intenzity a asynchronicity v rámci rozsahů daných trénovací množinou.



Obr. 24: Příklad uspořádání komponenty s parametry synchronní úlohy (vlevo) a asynchronní úlohy (vpravo).

Predikce komfortu je realizována na základě následujících vstupních dat: typu úlohy, typu proudění, šířky koridoru, intenzity (případně asynchronicity), označení profilu. Uvnitř systému jsou všechna tato vstupní data reprezentována maticí o velikosti 21x10, která je po zadání vstupních hodnot přeškálována a vstupuje do ANN. Vstupní matice X je společně s výstupní maticí Y zobrazena na Obr. 25. V prvních 15 sloupcích označených písmenem „a“ je zakódována kategorická proměnná do numerického vektoru reprezentující typ úlohy a proudění (např. „I\_uni“). Následuje sloupec „b“ reprezentující zadanou šířku profilu, dva sloupce „c“ reprezentující intenzity jednotlivých proudů (v případě jednosměrného proudění je druhý sloupec roven nule), sloupec „d“ značí asynchronicitu (v případě synchronních úloh je roven nule), sloupec „e“ značí vybraný profil a sloupec „f“ značí konkrétní bod na vybraném profilu. Na základě výstupní matice je následně vytvořen graf predikce komfortu, který je zobrazen v pravé části hlavního okna.

$$X = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,15} & b_{1,1} & c_{1,1} & c_{1,2} & d_{1,1} & e_{1,1} & f_{1,1} \\ a_{2,1} & a_{2,2} & \dots & a_{2,15} & b_{2,1} & c_{2,1} & c_{2,2} & d_{2,1} & e_{2,1} & f_{2,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{10,1} & a_{10,2} & \dots & a_{10,15} & b_{10,1} & c_{10,1} & c_{10,2} & d_{10,1} & e_{10,1} & f_{10,1} \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{10} \end{pmatrix}$$

Obr. 25 – Formát vstupní matice do ANN a výstupní matice.

Vstupní data specifikující typ úlohy a typ proudění jsou zadávána přepínači. Šířka koridoru a intenzita jsou řešeny posuvníky s předepsaným krokem. Zároveň je u těchto dat počítáno s možnou interpolací, tedy se zadáním šířky a intenzity, na které nebyla ANN přímo trénována, ale je schopna na jejich základě komfort predikovat. Dále je potřeba vymežit možný rozsah zadávaných dat, aby nedocházelo k extrapolaci, při níž je riziko rapidního snížení přesnosti predikce. Omezení rozsahu může být realizováno např. vyskakovacími okny, která uživatele upozorní na chybné zadání nebo přímým navedením uživatele k zadání správných parametrů. Druhá z možností je popsána výše a na Obr. 24.

## 7 ZÁVĚR

Tato metodika se zabývá návrhem, testováním a validací časoprostorovou predikcí komfortu osob a komplexní přípravou trénovacích dat generovaných z numerických modelů pohybu osob. Přesto nastíněné metody je možné užít i pro zpracování a predikce obecného charakteru – nejen nutně vázané na predikci komfortu, ale mohou mít přesah i mimo tento obor. Čtenář by si měl odnést postup, jak vytvořit základní úlohy pro predikce komfortu pohybu osob v budovách, teorii tvorby agentních simulací, možnou definici komfortu, přípravu dat pro metody strojového učení a základní informace pro výběr a provedení vhodné metody. Na závěr je uveden možný návrh softwarového nástroje sloužícího jako uživatelské rozhraní. Součástí metodiky není podrobný návod k vytvoření numerických simulací pohybu osob ani hlubší náhled do fungování umělých neuronových sítí nebo dalších metod strojového učení.

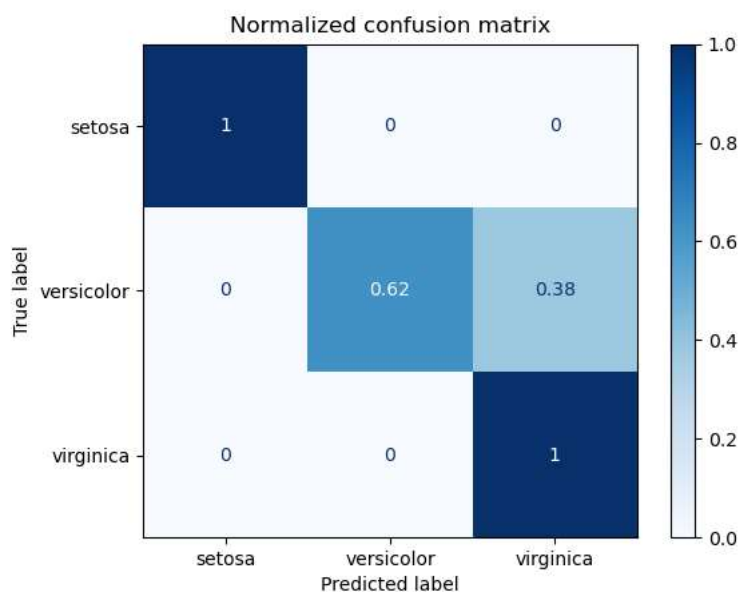
## PŘÍLOHY

### PŘÍLOHA A – METRIKY VYHODNOCENÍ

V případě, že by byl problém řešený modelem strojového učení definovaný jako klasifikační (nikoli regresní), jsou běžně využívány tyto metriky:

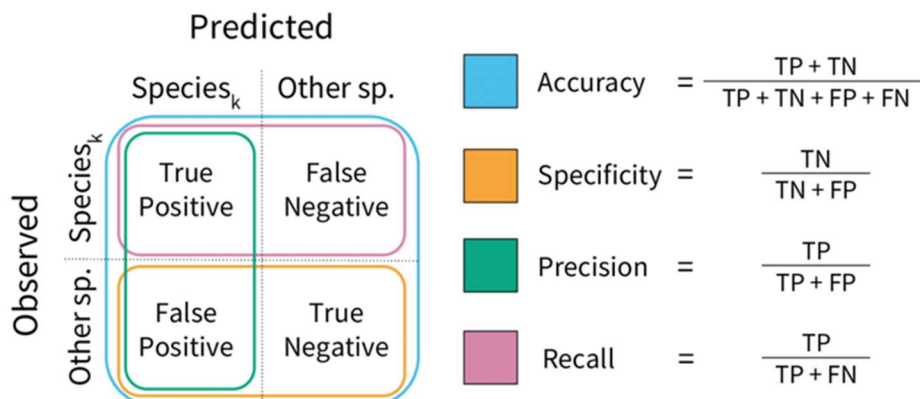
**Klasifikace** se zabývá rozdělením vstupních dat do předem stanovených tříd. V binární klasifikaci existují pouze dvě možné výstupní třídy (dichotomie). Ve vícetřídní klasifikaci jsou přítomny více než dvě **možné třídy**.

1. **Matice záměn** – Slouží k vizualizaci predikcí modelu. Každý řádek matice označuje instance v predikované třídě a každý sloupec označuje instance ve skutečné třídě (nebo naopak). Název matice je převzat ze skutečnosti, že umožňuje uživateli zkontrolovat, zda systém zaměňuje (tj. nesprávně klasifikuje) dvě třídy. Jedná se o konkrétní typ kontingenční tabulky, který má dvě dimenze (skutečnou a předpovězenou dimenzi) a stejné třídy v každé dimenzi (každá kombinace dimenze a třídy je identifikována jako proměnná).



Obr. 26 - Příklad matice záměn 39[19]

2. **Přesnost (Accuracy)** - hodnota kvantifikující, jak často klasifikátor správně předpovídá. Přesnost můžeme definovat jako poměr počtu správných předpovědí a celkového počtu předpovědí.
3. **Precision** – hodnota kvantifikující, kolik vzorků označených jako pozitivní je skutečně pozitivní.
4. **Recall** – kolik pozitivních vzorků bylo správně podchyceno klasifikátorem.



Obr. 27 - Metriky vycházející z matice záměn [20]

Různé metriky uvažující závažnost chyby 1. a 2. druhu

5. **F1 Score** – kombinace metriky precision a recall. Je maximální, když se metriky rovnají.
6. **AUC-ROC** – pro binární klasifikátory je další možností metriky ROC křivka neboli operační charakteristika přijímače (ROC = Receiver Operator Characteristic). Je to pravděpodobnostní křivka, která vykresluje pravděpodobnost detekce (True Positive Rate = Recall) proti pravděpodobnost falešného poplachu (False Positive Rate) pro různé klasifikační prahy. Klasifikační práh je hodnota pravděpodobnosti, od které se rozhodneme klasifikovat vzorek do jedné nebo druhé skupiny. Pokud musíme informace z grafu vyjádřit jedním číslem, používá se oblast pod křivkou (AUC).

## REFERENCE

---

- [1] NEUFERT, Ernst. Navrhování staveb: 1. vydání. Praha: CONSULTINVEST, 1995. 581 s.
- [2] Vyhláška č. 398/2009 Sb.: Vyhláška o obecných technických požadavcích zabezpečujících bezbariérové užívání staveb. In: 11/2009.
- [3] ČSN 73 0802. *Požární bezpečnost staveb: – Nevýrobní objekty*. 3. vyd. Praha: Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2020.
- [4] Pathfinder Technical Reference [online]. 403 Poyntz Avenue, Suite B, Manhattan, KS 66502, USA: Thunderhead Engineering, 2020 [cit. 2020-12-21]. Dostupné z: [https://www.thunderheadeng.com/wp-content/uploads/dlm\\_uploads/2011/07/tech\\_ref-6.pdf](https://www.thunderheadeng.com/wp-content/uploads/dlm_uploads/2011/07/tech_ref-6.pdf)
- [5] Pathfinder Verification and Validation [online]. 12. 403 Poyntz Avenue, Suite B, Manhattan, KS 66502, USA: Thunderhead Engineering: Thunderhead Engineering, 2020 [cit. 2020-02-27]. Dostupné z: <https://www.thunderheadeng.com/pathfinder-verification-validation-archive/>
- [6] Český statistický úřad – data o demografii v na území ČR 2020: *Ženy a muži v krajích ČR – Demografie* [online]. Praha: Český statistický úřad, 2020 [cit. 2022-05-08]. Dostupné z: [https://www.czso.cz/csu/xp/analyza-\\_zeny\\_a\\_muži\\_v\\_krajích\\_cr\\_demografie](https://www.czso.cz/csu/xp/analyza-_zeny_a_muži_v_krajích_cr_demografie)
- [7] *IMPVIS: Learn/Teach/Code* [online]. Level 5 Sherfield Building Imperial College London South Kensington Campus London SW7 2AZ (UK): Sherfield Building Imperial College London, 2021 [cit. 2022-05-08]. Dostupné z: <https://impvis.co.uk/launch/hydraulic-model/indexplaceholder.html>
- [8] FRUIN, John J. Pedestrian planning and design. New York: Metropolitan Association of Urban Designers and Environmental Planners, [1971]. ISBN 9781886536012
- [9] STEFFEN, Bernhard; SEYFRIED, Armin. Methods for measuring pedestrian density, flow, speed and direction with minimal scatter. *Physica A: Statistical mechanics and its applications*, 2010, 389.9: 1902-1910.
- [10] BUKÁČEK, Marek and Jana VACKOVÁ. Evaluation of pedestrian density distribution with respect to the velocity response. In *Traffic and Granular Flow'17*. Springer, 2019.
- [11] SCHADSCHNEIDER, Andreas, MOHCINE Chraibi, Armin SEYFRIED, Antoine TORDEUX, and Jun ZHANG, Pedestrian dynamics: From empirical results to modeling. In: *Crowd Dynamics, Volume 1*. Birkhäuser, Cham, 2018. p. 63-102.
- [12] HALL, Edward Twitchell; HALL, T. *The silent language*. Anchor books, 1959.
- [13] HALL, Edward Twitchell. *The hidden dimension*. Garden City, NY: Doubleday, 1966.
- [14] OKŘÍNOVÁ, Petra. *Využití modelů evakuace osob a šíření požáru pro zvýšení bezpečnosti specifických staveb*. Brno, 2022. Disertační práce. Vysoké učení technické v Brně, Fakulta stavební. Vedoucí práce Doc. Mgr. Tomáš Apeltauer, Ph.D.

- [15] SHIN, Jong Won; CHANG, Joon-Hyuk; KIM, Nam Soo. Voice activity detection based on statistical models and machine learning approaches. *Computer Speech & Language*, 2010, 24.3: 515-530.
- [16] KE, Shikun; OLEA, José Luis Montiel; NESBIT, James. A Robust Machine Learning Algorithm for Text Analysis. Working paper, 2019.
- [17] HOLČÍK, Jiří, KOMENDA, Martin (eds.) a kol. *Matematická biologie: e-learningová učebnice* [online]. 1. vydání. Brno: Masarykova univerzita, 2015. ISBN 978-80-210-8095-9.
- [18] MALATO, Gianluca. Yourdatateacher.com: Online courses and lessons about data science, machine learning and artificial intelligence. *Hyperparameter tuning. Grid search and random search* [online]. Via Roberto Malatesta 86, 00176 Roma (Italy): yourdatateacher.com, 2021, 2021 [cit. 2022-05-08]. Dostupné z: <https://www.yourdatateacher.com>
- [19] *Confusion matrix* [online]. scikit-learn developers, 2022 [cit. 2022-05-08]. Dostupné z: [https://scikit-learn.org/stable/auto\\_examples/model\\_selection/plot\\_confusion\\_matrix.html](https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html)
- [20] ANDERSON, Christopher B. The CCB-ID approach to tree species mapping with airborne imaging spectroscopy. *PeerJ*, 2018, 6: e5666.
- [21] Vacková Jana, Bukáček Marek. Kernel Estimates as General Concept for the Measuring of Pedestrian Density. <https://doi.org/10.48550/arxiv.2205.10145>, DOI 10.48550/ARXIV.2205.10145, arXiv 2022.
- [22] Vacková Jana, Bukáček Marek. Pedestrian Density Estimates and Their Real Applications. Stochastic and Physical Monitoring Systems, 2021.
- [23] Vacková Jana, Bukáček Marek. The Microscopic Analysis of Velocity-Density Paradigm. International Conference on Applied Mathematics (APLIMAT), 2019.
- [24] Vacková Jana, Bukáček Marek. Follower-Leader Concept in Microscopic Analysis of Pedestrian Movement in a Crowd. Pedestrian and Evacuation Dynamics 2018, Collective dynamics, 2020.