



**ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE**

**Fakulta elektrotechnická**

**Katedra telekomunikační techniky**

**RF navigace v LPWAN sítích**

**LPWAN RF Navigation**

Diplomová práce

Studijní program: MP77 – Elektronika a komunikace

Studijní obor: MPEK418 – Technologie internetu věcí

Vedoucí práce: doc. Ing. Stanislav Vítek, PhD.

**Bc. Tomáš Ouředník**

**Praha 2021**





# ZADÁNÍ DIPLOMOVÉ PRÁCE

## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Ouředník** Jméno: **Tomáš** Osobní číslo: **434767**  
Fakulta/ústav: **Fakulta elektrotechnická**  
Zadávající katedra/ústav: **Katedra telekomunikační techniky**  
Studijní program: **Elektronika a komunikace**  
Specializace: **Technologie internetu věcí**

## II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

**RF navigace v LPWAN sítích**

Název diplomové práce anglicky:

**LPWAN RF Navigation**

Pokyny pro vypracování:

1. Seznamte se s principy šíření RF signálu ve vybraných LPWAN sítích, zejména v sítích LoRa a SigFox.
2. Seznamte se s nejčastěji používanými metodami RF navigace, zejména AOA (Angle of Arrival), RSSI (Received Signal Strength Indication), TOA (Time of Arrival) a TR (Time Reversal).
3. Vybrané metody implementujte a pomocí vhodné platformy (např. SDR-LTR) experimentálně ověřte.
4. Diskutujte dosažené výsledky.

Seznam doporučené literatury:

- [1] GU, Chaojie; JIANG, Linshan; TAN, Rui. Lora-based localization: Opportunities and challenges. arXiv preprint arXiv:1812.11481, 2018.  
[2] LI, You, et al. Location-Enabled IoT (LE-IoT): A Survey of Positioning Techniques, Error Sources, and Mitigation. arXiv preprint arXiv:2004.03738, 2020.  
[3] MANZONI, Pietro, et al. Indoor vehicles geolocalization using LoRaWAN. Future Internet, 2019, 11.6: 124.

Jméno a pracoviště vedoucí(ho) diplomové práce:

**doc. Ing. Stanislav Vitek, Ph.D., katedra radioelektroniky FEL**

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **17.09.2020**

Termín odevzdání diplomové práce: **21.05.2021**

Platnost zadání diplomové práce: **19.02.2022**

doc. Ing. Stanislav Vitek, Ph.D.  
podpis vedoucí(ho) práce

podpis vedoucí(ho) ústavu/katedry

prof. Mgr. Petr Páta, Ph.D.  
podpis děkana(ky)

## III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

Datum převzetí zadání

Podpis studenta



### **Čestné prohlášení**

"Prohlašuji, že jsem předloženou práci vypracovala samostatně a že jsem uvedla veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací."

.....

datum

.....

podpis



**Abstrakt:**

Cílem práce je ověřit možnosti cenově a implementačně nenáročné techniky pro lokalizaci LPWAN zařízení v rámci zájmových budov. Pro tento účel byl navržen algoritmus využívající strojové učení a data o signálu Wi-Fi. Navržený algoritmus byl otestován se čtyřmi vybranými metody strojového učení a jejich vlastnosti byly v práci ověřeny pomocí dvou experimentálních měření.

**Klíčová slova:**

Lokalizace ve vnitřních prostorech, RSSI, LPWAN, strojové učení

**Abstract:**

The aim of this thesis is to verify the possibilities of cost-effective and low-cost implementation technology for localization of LPWAN equipment within buildings of interest. For this purpose, an algorithm using machine learning and data of Wi-Fi signal was designed. The proposed algorithm was tested with four selected machine learning methods and their properties were verified in the work using two experimental measurements.

**Key words:**

Indoor localization, RSSI, LPWAN, machine learning





# Obsah

---

1. Úvod .....	1
2. Internet věcí .....	2
3. LPWAN a Wi-Fi sítě.....	4
3.1 SigFox .....	6
3.1.1 Specifikace SigFox .....	6
3.2 LoRa.....	9
3.2.1 Specifikace LoRa .....	10
3.3 Wi-Fi .....	13
3.3.1 Specifikace Wi-Fi .....	14
4. Strojové učení .....	15
4.1 Základní typy strojového učení .....	15
4.2 Typy strojového učení s dohledem .....	16
4.3 Vyhodnocení modelu strojového učení .....	19
5. Rozbor RF lokalizačních systémů .....	21
5.1 Systémy založené na modelu šíření signálu .....	21
5.2 Přístupy založené na otiscích signálu .....	23
5.3 Vlastnosti lokalizačních systémů .....	24
5.4 Současně oficiálně nabízené lokalizační služby v LPWAN sítích .....	28
6. Návrh a implementace algoritmu .....	29
6.1 Implementace .....	29
6.2 Struktura algoritmu .....	31
7. Testování .....	32
7.1 Testovací scénář A.....	34
7.2 Testovací scénář B.....	41
8. Závěr.....	44
Příloha A.....	46
Reference .....	49

# Seznam použitých obrázků

---

Obr. 1 Vrstvy IoT.....	2
Obr. 2 Porovnání základních typů bezdrátových sítí.....	4
Obr. 3 Naměřené spektrum základnové stanice SigFox Pankrác, Praha.....	5
Obr. 4 Naměřené spektrum základnové stanice SigFox Kublov, okres Beroun .....	5
Obr. 5 Pokrytí sítě SigFox .....	6
Obr. 6 Rozdělení pásma SigFox .....	7
Obr. 7 D-BPSK modulace .....	7
Obr. 8 Frekvenční skákání .....	8
Obr. 9 Kooperativní příjem zpráv .....	8
Obr. 10 Mapa pokrytí LoRa v České republice .....	9
Obr. 11 Architektura LoRaWAN .....	10
Obr. 12 Zprávy sítě LoRa zachycené na základnové stanici SigFox .....	11
Obr. 13 Módy komunikace LoRa a jejich specifikace .....	12
Obr. 14 Porovnání Wi-Fi standardů.....	13
Obr. 15 Rozdělení pásma 2,4 GHz na jednotlivé kanály.....	14
Obr. 16 Jednoduché vysvětlení KNN algoritmu .....	16
Obr. 17 Příklad rozdělující roviny SVM ve dvou a tří rozměrném prostoru.....	17
Obr. 18 Schéma algoritmu RF.....	18
Obr. 19 Schéma K-násobné křížové validace .....	19
Obr. 20 Schéma Monte Carlo křížové validace .....	20
Obr. 21 Porovnání technik pro výpočet vzdálenosti .....	21
Obr. 22 Schéma triangulace .....	22
Obr. 23 Off-line fáze .....	23
Obr. 24 On-line fáze .....	24
Obr. 25 Závislosti úspěšnosti určení polohy na šířce pásma. a) s pohybem osob, b) s pohybem velkých předmětů.....	27
Obr. 26 Diagram funkce learn .....	31
Obr. 27 Diagram funkce train.....	31
Obr. 28 Topologie testovací lokality.....	32
Obr. 29 Rozložení referenčních bodů pro jednotlivé místnosti .....	34
Obr. 30 Rozložení jednotlivých referenčních bodů v místnosti č. 2.....	41

# 1. Úvod

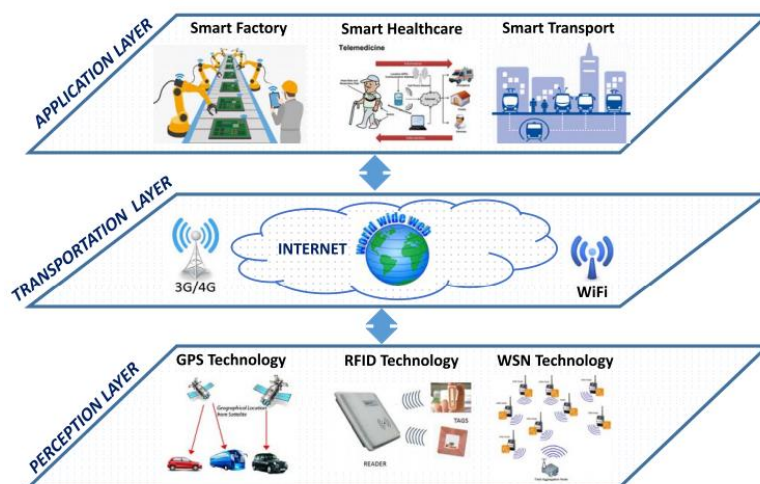
---

V dnešní době se pro komerční a průmyslové použití prosazují lokalizační technologie ve vnitřních prostorech. Mezi hlavní oblasti aplikace patří například trasování zboží, správa tras nebo navigace v komplexních budovách, jako jsou letiště, výstavní haly a nákupní centra. Tyto technologie mají potenciál rozšířit funkci inteligentních budov, zlepšit navigaci a snížit časovou náročnost při hledání obchodů či zboží. Provozovatelům nákupních center umožňuje poskytovat obsah a personalizovanou reklamu díky sledování pohybu zákazníků, nebo pomocí tzv. Bluetooth majáků zasílat oznámení o slevových kuponech a speciálních nabídkách v okolí zákazníků. Na letištích mohou uživatelé s aplikací v telefonu použít navigaci a případně zjistit, kde jsou kontrolní přepážky a jejich vytíženost, vzdálenost k bráně, čas nástupu a podobně. Sledování věcí v reálném čase usnadňuje vyhledání zavazadel, invalidních vozíků a dalších věcí. V muzeích lokalizace umožňuje navigovat návštěvníky různými odděleními a doplnit prohlídku o audio průvodce nebo obsah spouštěný polohou. Kvůli omezení globálního polohového systému (GPS) pro použití ve vnitřních prostorech [1] se objevují služby založené na jiných rádiových protokolech. V současné době se prosazují, zejména kvůli cenové dostupnosti, protokoly jako jsou Wi-Fi, Bluetooth nebo RFID.

Úvodu do problematiky internetu věcí s tím spojených LPWAN sítí a Wi-Fi jsou věnovány kapitoly dva a tři. Jelikož navrhovaný algoritmus využívá pro lokalizaci strojové učení, jsou ve čtvrté kapitole rozebrány základní typy modelů strojového učení a metody ověření jejich přesnosti. Rozboru technologií pro lokalizaci ve vnitřních prostorech a jejich vlastnostem je věnována pátá kapitola. V šesté a sedmé kapitole je navržen a otestován algoritmus pro lokalizaci zařízení ve vnitřních prostorech.

## 2. Internet věcí

IoT (Internet of Things) je koncept, ve kterém je většina běžných zařízení, v domácnosti nebo průmyslu, připojena k internetu a autonomně komunikuje s ostatními zařízeními. IoT má potenciál přinést nové obchodní příležitosti a měnit dynamiku průmyslu [2]. V minulosti bylo několik odhadů trendu vývoje růstu počtu zařízení. Společnost CISCO v roce 2011 odhadovala, že do roku 2020 bude připojeno 50 miliard IoT zařízení [3], společnost zabývající se technologickým výzkumem Gartner v roce 2013 odhadovala 26 miliard IoT zařízení [4] do roku 2020. V současné době z veřejně dostupných materiálů není možné určit číslo komunikujících IoT zařízení přesně. Web Statista odhaduje, že na konci roku 2020 bylo připojeno 190 milionů zařízení na síti LoRa, 130 milionů na síti NB-IoT a 20 milionů na síti SigFox [5]. Ekosystém IoT se typicky skládá ze 3 samostatných vrstev, a to senzorové, transportní a aplikační [6].



Obr. 1 Vrstvy IoT [6]

Do senzorové vrstvy patří například senzory veličin umístěné v objektech nebo nositelné elektronice. Podle typu aplikace může jít například o senzory CO<sub>2</sub>, osvětlení, průtoku kapalin, naplnění popelnic, GNSS atd. Senzory pro IoT čelí několika výzvám a to:

- Nízká cena
- Dlouhá životnost
- Komunikace na velkou vzdálenost
- Vysoká hustota a škálovatelnost

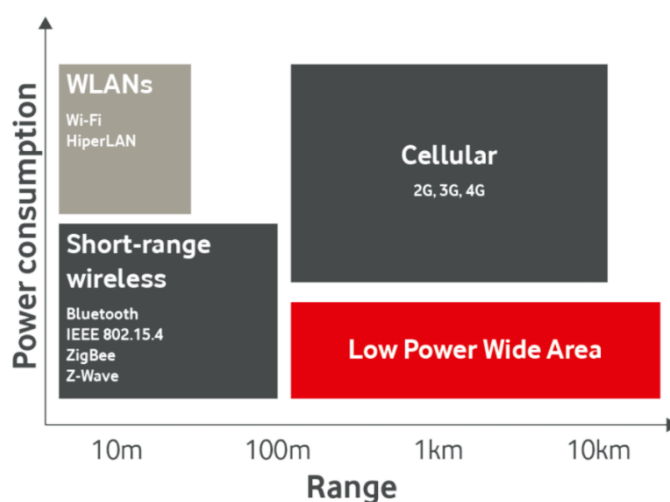
Transportní vrstva se zaměřuje na komunikaci typu end-to-end. Jako vlastnosti, které ji definují může být uvedena spolehlivost doručení zprávy a odolnost vůči přetížení. Účelem této vrstvy je přenášení shromážděných informací přijatých senzory do centrálního prvku, který zajišťuje zpracování informací prostřednictvím stávajících komunikačních sítí jako je například 3G, 4G, Wi-Fi nebo ad hoc sítí.

Aplikační vrstva zpracovává a vizualizuje hrubá data ze sensorů podle požadavků zákazníků, například převod zprávy o naměřené teplotě a vlhkosti vzduchu ze senzoru do přehlednější formy zobrazení jako jsou grafy. Na této úrovni lze implementovat nepřeberné množství aplikací IoT. Navíc aplikace mohou být propojeny do složitějších celků, které mohou pomoci optimalizovat procesy v průmyslu, logistice, nebo zdravotnictví. Například propojením aplikací na sledování vytížení dopravních spojů, zaplněnosti odpadních kontejnerů a recyklačních center může být optimalizován svoz odpadu a zpracování odpadů.

### 3. LPWAN a Wi-Fi sítě

S rozšířením IoT (Internet of Things) se v posledních letech na trhu objevují zařízení komunikující přes LPWAN sítě (Low-Power Wide-Area Network). Jedná se o sítě, které umožňují přenos menšího objemu dat, desítky kilobajtů až nižší jednotky megabajtů za měsíc, na dlouhé vzdálenosti s vysokou energetickou účinností. Díky tomu mohou IoT aplikace přenášející data přes LPWAN využívat senzory, které jsou napájené z baterie, i deset let. Porovnání základních typů bezdrátových sítí je na Obr. 2.

Vysoká energetická účinnost je dosahována různými způsoby. LoRa (Long Range) využívá modulaci rozprostřeného spektra s volitelným poměrem počtu přenášených bitů a citlivostí radiového spoje. Díky tomu může být snížením počtu přenášených bitů zvýšena citlivost radiového spoje. SigFox využívá UNB (ultra-narrow band) pásma s vysokou energetickou hustotou [7].

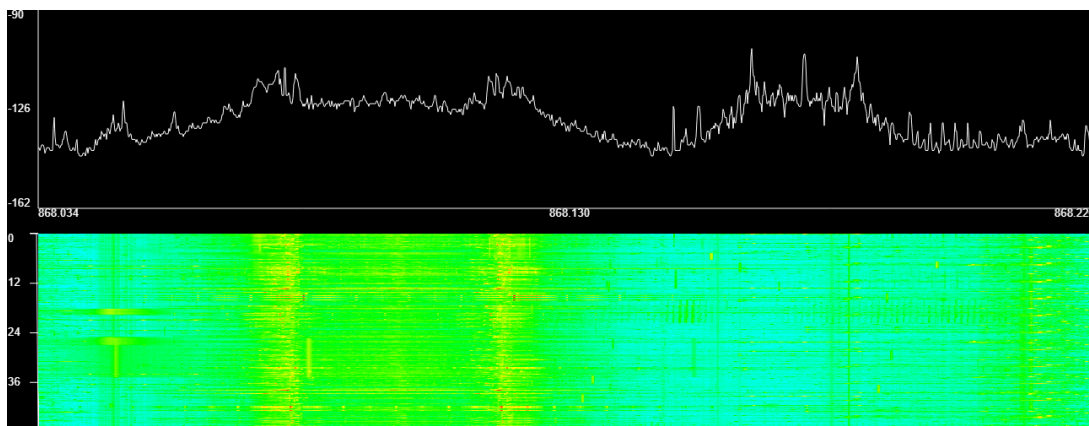


Obr. 2 Porovnání základních typů bezdrátových sítí

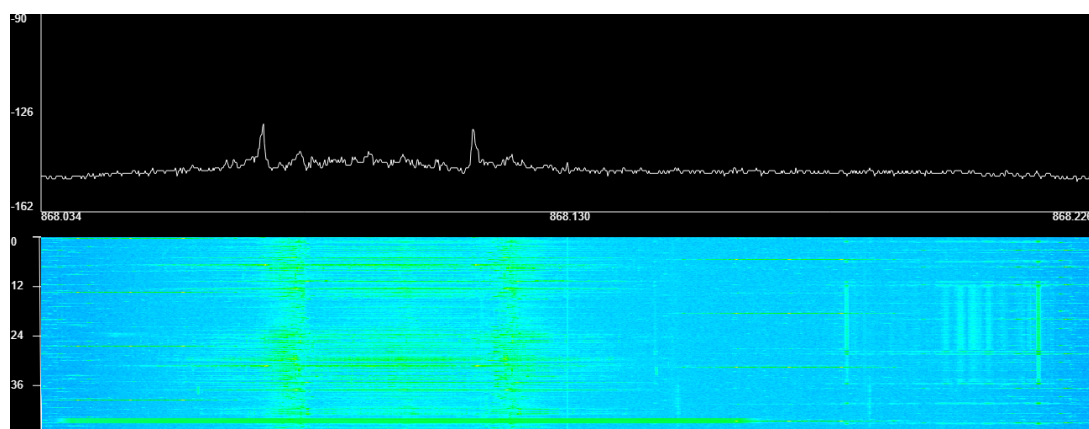
LPWAN sítě typicky pracují na bezlicenčních rádiových frekvencích, na území Evropské Unie například 800-900 MHz nebo 2,4GHz. LoRa a SigFox v EU využívá nosnou frekvenci 868 MHz, ve Spojených státech amerických pak frekvence od 902 do 928 MHz. Zigbee vysílá nejčastěji na 2,4GHz. Nižší frekvence má určité výhody – nižší energetická náročnost a útlum volného prostoru nebo snazší prostup signálu překážkami. Nicméně nižší frekvence má své praktické limity ve velikosti antény, jelikož se zvyšující se délkou vlny je potřeba větší anténa. Na tato bezlicenční pásma se vztahuje evropská regulace, podle které zařízení vysílající na frekvencích v těchto pásmech mohou vysílat s maximálním výkonem 25 mW, a to po dobu 1% nebo 0,1%

vysílací hodiny, v závislosti na použité frekvenci [8]. Využití bezlicenčního pásma přináší velkou finanční úsporu, jelikož operátoři nemusí platit licenční poplatky za využívané přenosové frekvence.

Na druhou stranu využívání bezlicenčního pásma přináší i nevýhody. Bezlicenční pásma využívají i velmi rozšířené technologie jako je Wi-Fi a Bluetooth. Bezlicenční pásmo kolem 868MHz se například v EU používá i pro dálkové ovládání semaforů a zařízení v domácnostech. To způsobuje vysoké rušení, obzvláště ve větších městech což je vidět na Obr. 3 a Obr. 4. Toto rušení způsobuje nižší citlivost základnové stanice a může mít za následek nízkou kvalitu služeb [9]. Například v aplikaci zdravotní péče založené na IoT musí být kritické informace o pacientovi předány lékaři, což v případě zvýšeného rušení způsobeného ostatními technologiemi nemusí být zaručeno.



Obr. 3 Naměřené spektrum základnové stanice SigFox Pankrác, Praha

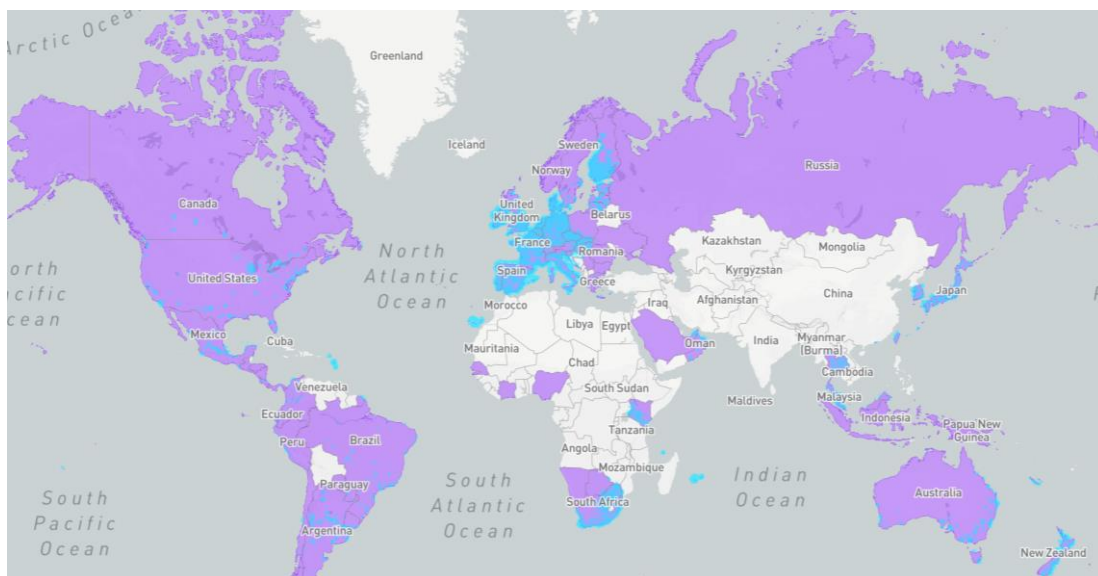


Obr. 4 Naměřené spektrum základnové stanice SigFox Kublov, okres Beroun

## 3.1 SigFox

SigFox je firma založená ve Francii v roce 2010, která vlastní stejnojmennou proprietární LPWAN technologii. Proprietární v tom smyslu, že firma vlastní celý ekosystém a tzn. SigFox operátorům, prodává kompletní infrastrukturu jako např. síťový hardware, přístup do Sigfox webového rozhraní a dalších služeb v cloudu.

V současné době SigFox působí ve více než 70 zemích, kde pokrývá 5,8 milionu km<sup>2</sup>. V síti komunikuje 17,2 milionu zařízení a za den zpracuje cca 63 milionu zpráv.

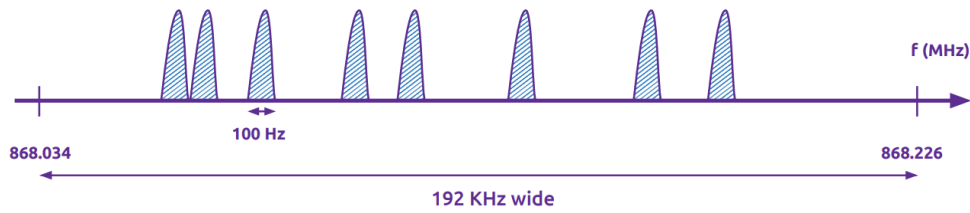


Obr. 5 Pokrytí sítě SigFox [10]

### 3.1.1 Specifikace SigFox

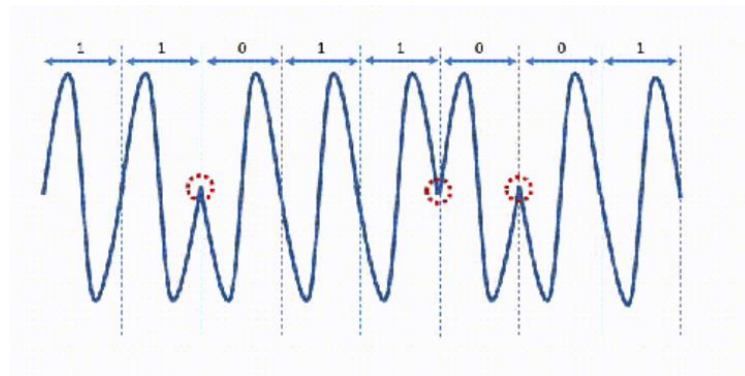
Fyzická vrstva SigFox využívá topologii hvězdy, kdy s jednou základnovou stanicí, která převádí SigFox zprávy na IP a přeposílá dále do SigFox backendu s využitím stávajících celulárních nebo satelitních sítí. S jednou základnovou stanicí mohou komunikovat i tisíce zařízení. Nicméně v praxi se, ale pro zvýšení spolehlivosti doručení zprávy, základové stanice staví tak, aby zprávu od zařízení přijaly vždy alespoň dvě základnové stanice. Zařízení komunikují v kanálu o šířce 100 Hz pro up-link a 600 Hz pro down-link. Takto široká pásma spadají do kategorie "ultra-narrow band", která se vyznačuje vysokou spektrální hustotou a selektivitou, dokáže potlačit šum a rušení, které mohou vstoupit do přijímače mimo jeho úzkou šířku pásma, což umožňuje dosáhnout přijatelného poměru signálu k šumu i s relativně slabým přijímaným signálem [11].





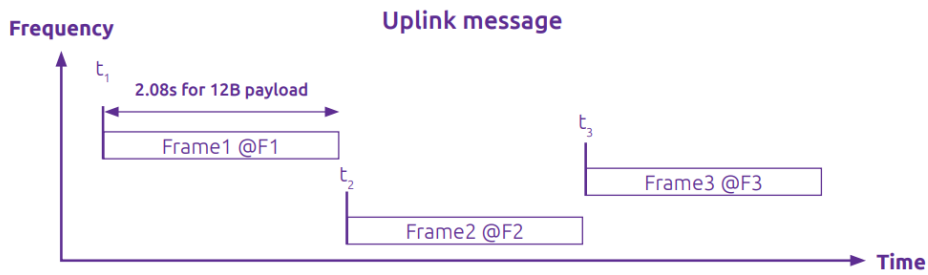
Obr. 6 Rozdělení pásma SigFox [11]

SigFox využívá modulaci D-BPSK (Differential Binary Phase-Shift Keying). Rychlost odesílání je 100 b/s v Evropě a 600 b/s v USA.



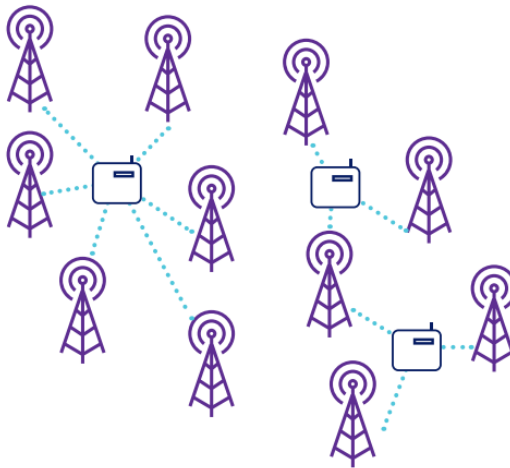
Obr. 7 D-BPSK modulace

Šumové dno základnová stanice je v ideálním případě, tzn. bez vnějšího rušení, -142 dBm. Podle evropské regulace bezlicenčních pásem mohou zařízení komunikující přes síť SigFox vysílat 1% hodiny, to je 36 sekund. Velikost zprávy může být až 26 bajtů z toho až 12 bajtů obsahuje uživatelská data. Každá zpráva se vysílá třikrát [11]. Rychlostí 100 b/s se zpráva odesílá 2,08 sekundy. Pro zprávu obsahující 12 bajtů to znamená 5,77 zpráv za hodinu, zhruba 138 zpráv denně pro up-link. Pro down-link SigFox umožňuje vysílat 4 zprávy denně, každou s 8 bajty dat. Z důvodu úspory energie, není mezi zařízeními a sítí žádná synchronizace. Základnová stanice SigFox musí být i bez synchronizace a centrálně řízeného provozu v síti schopna odbavit zprávy od tisíců zařízení denně. Zabránění kolize zpráv při sdíleném pásmu všech komunikujících zařízení při příjmu zprávy je docíleno časovou a frekvenční diverzitou společně s prostorovou diverzitou [11].



Obr. 8 Frekvenční skákání [11]

Pro vysílání zprávy zařízení zvolí náhodnou frekvenci v pásmu, poté zařízení vyšle dvě kopie zprávy na jiných frekvencích a náhodným časovým odstupem. To znamená, že se každá taková zpráva vysílá zhruba 6,24 sekund. Toto je nazýváno jako časová a frekvenční diverzita.



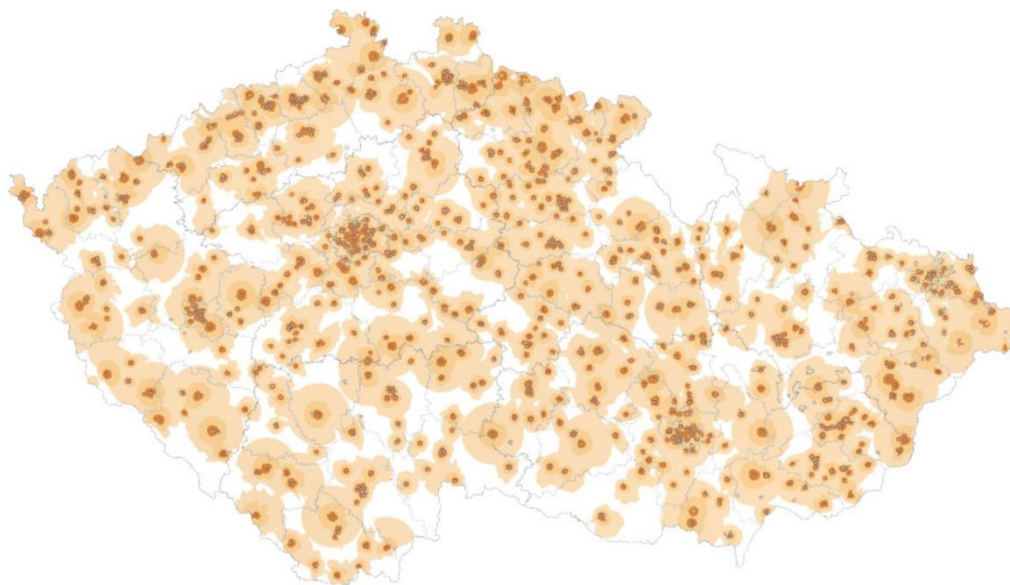
Obr. 9 Kooperativní příjem zpráv [11]

Narozdíl od buňkových bezdrátových technologií jednotlivá zařízení v síti SigFox nejsou připojena pouze k jedné specifické základnové stanici. Vysílaná zpráva je přijata všemi základnovými stanicemi v dosahu. Průměrný počet základnových stanic, které přijímají zprávu v síti SigFox, je tři [11]. Příjem jedné zprávy na různě vzdálených základnových stanicích je nazýváno prostorovou diverzitou.

## 3.2 LoRa

LoRa a LoRaWAN vznikala od 2009 ve Francii ve firmě Cycleo, kterou v roce 2021 odkoupila a patentovala firma Semtech, která v současné době prodává LoRa čipy. LoRa je patentovaná technologie, která moduluje data na fyzické vrstvě tak, aby umožnila komunikaci na velké vzdálenosti. LoRaWAN definuje protokol přístupové vrstvy pro řízení komunikace mezi bránami LPWAN a koncovým zařízením. Specifikace LoRaWAN je definována LoRa aliancí, verze 1.0 byla vydána v roce 2015. Aliance LoRa pro zařízení poskytuje certifikaci, která dokládá, že zařízení je kompatibilní s LoRaWAN a splňuje regulaci v daném státu. Mezi zakládající členy aliance LoRa patří například Cisco, IBM, Kerlink, Semtech nebo Microchip Technology. Na rozdíl od technologie SigFox je LoRa volně přístupná. Díky tomu na ní ostatní společnosti mohou stavět vlastní řešení bez nutnosti platit licenční poplatky a pro výstavbu mohou použít zařízení od různých výrobců.

V České republice si technologii LoRa pro budování sítě IoT vybrala firma CRA (České Radiokomunikace). Pro připojení zařízení do sítě LoRa CRA je nutné, aby zařízení od CRA dostalo certifikaci. To nutí výrobce IoT zařízení dívat se na IoT i z bezpečnostního hlediska.

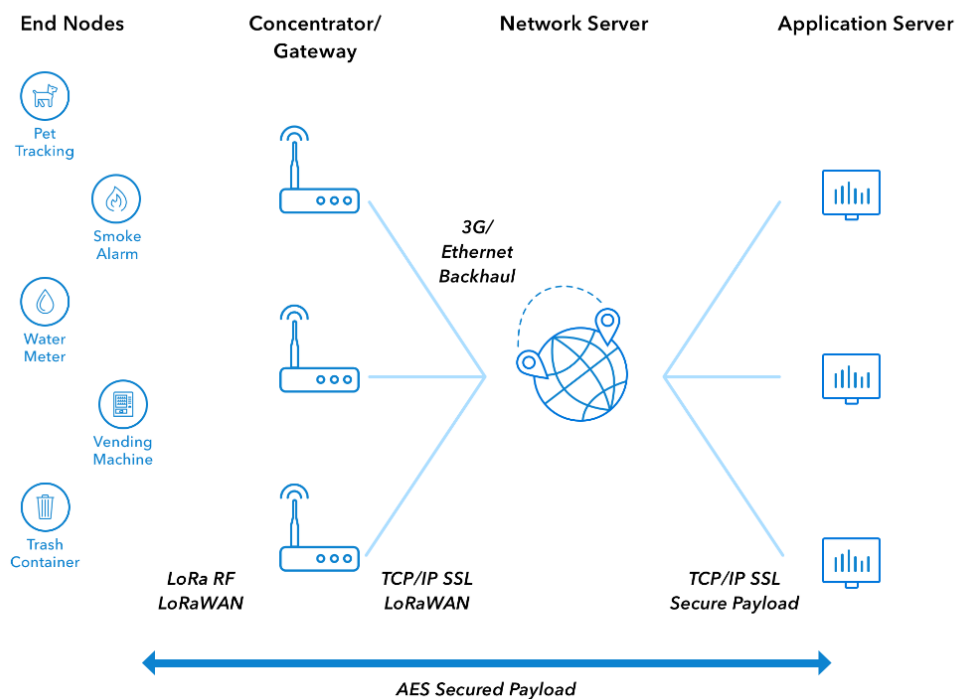


Obr. 10 Mapa pokrytí LoRa v České republice [12]

### 3.2.1 Specifikace LoRa

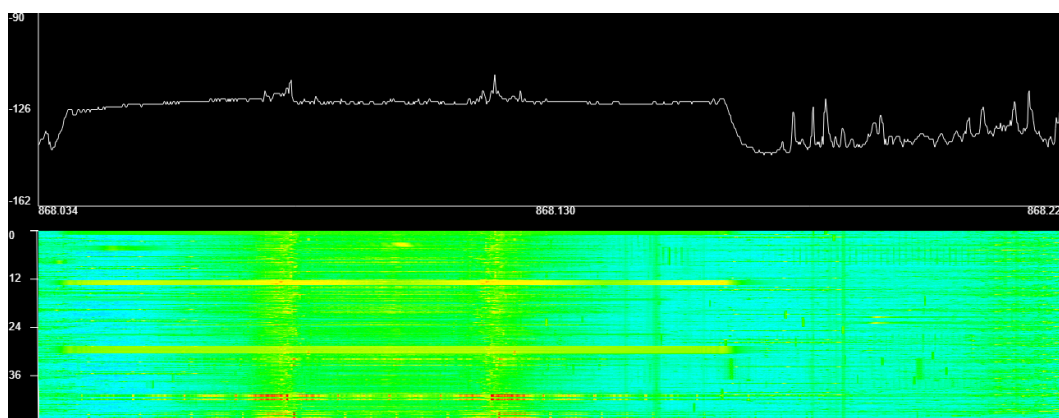
Topologie LoRaWAN je obvykle hvězdicová a její architektura se skládá ze 4 vrstev, a to:

- Koncové zařízení – Například senzory nebo akční členy, typicky napájené baterií. Zařízení v síti vysílají buď periodicky, nebo jakmile jsou k dispozici data.
- LoRa brána (gateway) – V LoRaWAN brány nahrazují základnové stanice používané u SigFox nebo celulárních sítích. Každá brána přijímá zprávy od všech zařízení v dosahu a přeposílá na centrální server typicky po IP síti jako například 3G/4G, Ethernet nebo satelitní spojení. Ke zprávám jsou na bráně připojeny data o přijetí, například časová značka, informace o kvalitě signálu přijímané zprávy nebo použité pásmo.
- Centrální server – Centrální server má na starosti správu jednotlivých bran, filtraci/agregaci duplicitních paketů přijatých více branami, řízení vysílání bran a kontrolu bezpečnosti. Navíc může řídit úroveň vysílacího výkonu koncových zařízení. Zpracovaná data poté přeposílá na aplikační server.
- Aplikační server – Aplikační server je typicky spravován zákazníkem, který pracuje s přijatými daty. Dále data mohou být analyzována a zpracována do statistik nebo grafů.



Obr. 11 Architektura LoRaWAN [13]

Základem fyzické vrstvy LoRa je modulační schéma využívající rozprostřené spektrum, které je založeno na CSS (Chirp Spread Spectrum) modulaci. Modulační schéma využívá širokopásmové lineárně frekvenčně modulované pulsy, ke kódování dat se používá zvýšení nebo snížení frekvence v čase. Pokud je modulace kombinována s dopřednou korekcí chyb, mohou být demodulovány signály, které jsou až 20 dB pod úrovní šumu a oproti tradičnímu modulování frekvenčním posuvem zvyšuje modulace LoRa citlivost o 25 dB [14]. LoRa zpráva zachycená na základnové stanici SigFox je na Obr. 12.



Obr. 12 Zprávy sítě LoRa zachycené na základnové stanici SigFox

V určité vzdálenosti nebo při vhodném nastavení vysílacího výkonu se signál LoRa zařízení ztrácí v šumu a mimo zařízení podporující modulaci LoRa jej může být obtížné detekovat. To může být s výhodou využito v zabezpečovacích systémech. Jako další výhodou této modulace může být uvedena odolnost vůči frekvenčním offsetům, díky čemuž může být pro zařízení využito méně přesný a tím pádem levnější krystalový oscilátor. Navíc na rozdíl od SigFox protokolu LoRa je výrazně odolnější vůči frekvenčnímu posuvu způsobeného pohybem zařízení [15].

LoRa umožňuje komunikaci oběma směry, tj. jak od zařízení k bráně, tak i naopak. Pro každý směr jsou využívána jiná frekvenční pásma a z důvodu optimalizace spotřeby energie i rozdílné přenosové rychlosti. Volbou přenosové rychlosti, tzn. přenosového módu, může být optimalizován poměr mezi komunikačním dosahem a dobou odesílání zprávy, tj. bitovou rychlost. Při komunikaci na větší vzdálenost se může snížit bitová rychlost a tím zvýšit pravděpodobnost přijetí zprávy. Takováto optimalizace umožňuje vytvoření virtuálních kanálů, které vzájemně neinterferují a zvyšují tak kapacitu brány nebo snižují spotřebu energie nutnou pro úspěšné přijetí zprávy. Specifikace módů pro komunikaci LoRa jsou na Obr. 13 Módy komunikace LoRa a jejich specifikace.

Mode	Bitrate (bits/sec)	Max payload size (bytes)
SF7/125kHz	5470	222
SF8/125kHz	3125	222
SF9/125kHz	1760	115
SF10/125kHz	980	51
SF11/125kHz	440	51
SF12/125kHz	250	51
SF7/250kHz	11000	222

Obr. 13 Módy komunikace LoRa a jejich specifikace [13]

### 3.3 Wi-Fi

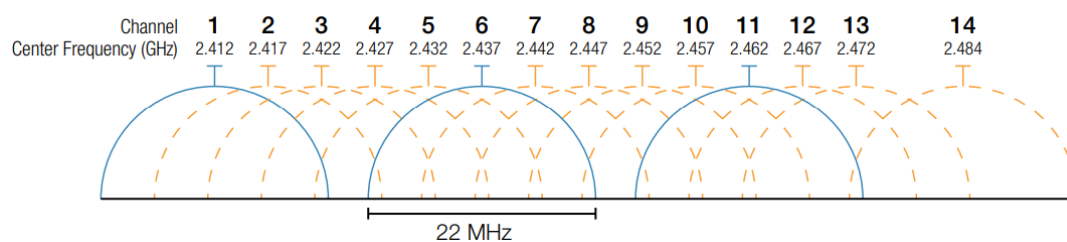
Wi-Fi je obchodní označení rodiny protokolů IEEE 802.11, které se ve velkém využívají pro lokální bezdrátové připojení zařízení a přístup k internetu. Tyto protokoly spadají pod neziskovou organizaci IEEE (Institute of Electrical and Electronics Engineers), IEEE 802 je skupina standardů zabývajících se místní (LAN) a metropolitní sítí (MAN). Kromě Wi-Fi do této skupiny patří například i 802.3 Ethernet, 802.15.1 Bluetooth a 802.16 WiMAX. První verze protokolu Wi-Fi byla uvedena roku 1997, roku 1999 byla poté založena aliance firem vyvíjejících a implementujících Wi-Fi s cílem zajistit kompatibilitu zařízení bez ohledu na výrobce. První verze Wi-Fi měla maximální přenosovou rychlost 2 Mbit/s, nicméně během následujících let bylo vydáno několik dalších verzí, které zvyšovali například přenosovou rychlost, spolehlivost nebo zabezpečení. Z Wi-Fi se během té doby stala jedna z dominantních bezdrátových technologií a bezdrátovou analogií ethernetu, počty zařízení stoupají stejně jako přenášené datové toky. Dle údajů Wi-Fi Alliance bylo v roce 2018 dodáno na trh okolo 3 miliard zařízení podporující komunikaci pomocí Wi-Fi [16].

Technologies	Indoor/ Outdoor	Bitrate	Freq. bands	License	Bandwidth	Modulation	MIMO
IEEE 802.11	20m /100m	2 Mbps	2.4GHz	Unlicensed	20 MHz	FHSS and DSSS	—
IEEE 802.11b	35m/ 140m	11 Mbps	2.4GHz	Unlicensed	20 MHz	HR-DSSS	—
IEEE 802.11a	35m/ 119m	54 Mbps	5GHz	Unlicensed	20 MHz	OFDM	—
IEEE 802.11g	45m/ 90m	54 Mbps	2.4 GHz	Unlicensed	22 MHz	OFDM/ DSSS/ CCK	—
IEEE 802.11n	70m/ 250m	600 Mbps	2.4 GHz/ 5 GHz	Unlicensed	20 MHz/ 40 MHz	OFDM	4 X 4
IEEE 802.11ac wave	70m/ 250m	7000 Mbps	5 GHz	Unlicensed	80 MHz	64-QAM	MU-MIMO
IEEE 802.11ad	10m/ n/a	7000 Mbps	60 GHz	Unlicensed	2.16 GHz	Single Carrier/ OFDM	10 X 10
IEEE 802.11ac wave 2	70m/ 250m	7000 Mbps	5 GHz	Unlicensed	80 MHz/ 160 MHz	256-QAM	MU_MIMO 8 X 8

Obr. 14 Porovnání Wi-Fi standardů [17]

### 3.3.1 Specifikace Wi-Fi

Wi-Fi využívá obvykle hvězdicovou topologii a bezlicenční pásma na frekvencích 2,4 a 5 GHz. Šířka pásma kanálu se liší podle nosné frekvence, na 2,4GHz je to až 40MHz a na 5GHz poté až 160MHz. Od nosné frekvence se odvíjí i další parametry Wi-Fi, a to například dosah, propustnost, kompatibilita a mezi-kanálové interference. Pásmo od 2,4000 do 2,4835 GHz je rozděleno do třinácti 22 MHz širokých kanálů, které jsou vzájemně posunuté o 5 MHz. V Japonsku se používá i čtrnáctý kanál, který je posunutý o 12 MHz od třináctého. Z toho plyne, že pokud nemá docházet k vzájemnému rušení, mohou běžet souběžně pouze tři Wi-Fi sítě například na kanálech 1, 6, 11 [18].



Obr. 15 Rozdělení pásma 2,4 GHz na jednotlivé kanály [18]

Pásmo na 5GHz je širší a to od 5,180 GHz do 5,700 GHz. V Evropě Wi-Fi může vysílat na 19 kanálech, z nichž je 8 určeno pouze pro použití uvnitř budov. Jednotlivé kanály na 5GHz pásmu jsou vzájemně posunuté o 20 MHz, při použití 20 MHz kanálu tak lze provozovat devatenáct 5GHz Wi-Fi sítí na jednom místě bez vzájemného rušení. Oproti pásmu 2,4 GHz má 5 GHz nižší dosah způsobený vyšší nosnou frekvencí, která má větší útlum při šíření a některá aktuálně prodávaná zařízení s tímto pásmem nejsou kompatibilní. Na druhou stranu umožňuje 5GHz pásmo vyšší přenosovou rychlost a z důvodu vyššího útlumu a širšího pásma je méně náchylné k rušení.

Síla signálu Wi-Fi může být hodnocena pomocí hodnoty RSSI (Received Signal Strength Indicator). V protokolu IEEE 802.11 může RSSI nabývat hodnot od 0 do 255. Výrobce čipové sady může definovat vlastní maximální hodnotu RSSI. Například Cisco používá stupnici 0-100, zatímco Atheros 0-60. Čím vyšší je hodnota RSSI, tím lepší je signál. Hodnota RSSI se měří v preambuli přijímacího paketu. Jakmile je detekována preambule, přijímač signalizuje vrstvě MAC aktivitu na detekovaném kanálu a předává naměřenou hodnotu RSSI. Čipová sada může použít metodu pro výpočet hodnoty RSSI vztahenou k SNR (Signal-to-noise ratio). Hodnota RSSI poté představuje počet dBm nad úrovní šumu [19].



## 4. Strojové učení

---

S rostoucím množstvím dat, která jsou denně generována rostou i nároky na jejich analýzu, právě z toho důvodu se strojové učení v posledních dvou desetiletích stalo jedním z hlavních směrů vývoje v informačních technologiích. Strojového učení spadá pod obor umělé inteligence a může být využito například k vyhledání vzorců v datech a předpovídání budoucích událostí. Základní algoritmus, který je naprogramován člověkem, se však místo přímého lidského zásahu zdokonaluje na základě předložených dat, tzv. trénovacích dat, a utvoří matematický model na základě statistiky. Model lze poté použít k analýze nových, tzv. testovacích dat. Algoritmy pro strojové učení je možné rozdělit podle několika parametrů. Základní dělení může být podle toho, jak algoritmy zachází se vstupními a výstupními daty.

### 4.1 Základní typy strojového učení

#### Supervised learning – („učení s dohledem“)

Algoritmus se učí na základě trénovacích dat, která se skládají ze vstupních parametrů a požadovaného výstupu. Cílem je vytvořit funkci neboli model, který pro libovolný vstup přiřadí správný výstup. Model musí zobecnit parametry modelu tak, aby se dal aplikovat i na jiná, dosud nepředložená data a vyhnout se pouhému zapamatování trénovacích dat. Takzvané super-vizované učení může být využito například pro regresi nebo klasifikaci. Regrese je metoda modelování budoucí hodnoty na základě nezávislých prediktorů [20]. Tato metoda se používá pro předvídaní a zjišťování vztahu příčin a následků mezi proměnnými, jejím výstupem bývá typicky číselná hodnota [21]. Cílem klasifikace je přiřadit vzorky na základě vstupních proměnných, typicky označovaných jako „X“, do jedné z daných kategorií, typicky označovaných jako „y“ [22].

#### Un-supervised learning – („učení bez dohledu“)

Na rozdíl od předchozího, učení bez dohledu nemá informaci o požadovaném výstupu ke vstupním parametrům. Typicky jediným vstupem je v tomto případě velké množství dat a parametrů pro jednotlivé vzorky. Cílem je najít vzory mezi těmito daty, k tomu se často používá měřítko podobnosti mezi vzorky dat nebo modelování jejich hustoty pravděpodobnosti [23].

## Reinforcement Learning – („Posílené učení“)

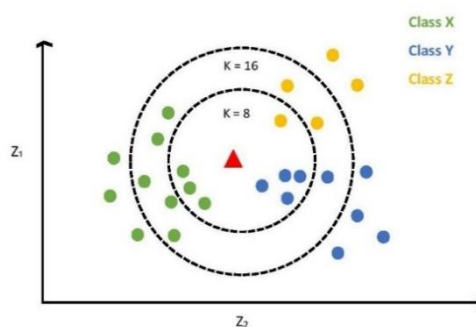
Posílené učení je typ strojového učení pravděpodobně nejbliže tomu, jak se učí lidé. Algoritmus se učí dosáhnout cíle v nejistém, potenciálně složitém prostředí a čelí problému podobnému hře, kterou se snaží vyřešit metodou pokus – omyl. Aby algoritmus dělat to, co programátor chce, dostává odměny nebo pokuty za akce, které provádí a jeho cílem je maximalizovat celkovou odměnu. Programátor tedy nastavuje pouze princip odměňování. Je na algoritmu, aby zjistil, jak úkol splnit a maximalizovat tak odměnu. Díky využití výpočetní síly, které dovoluje algoritmu provádět tisíce paralelních pokusů, je tento typ strojového učení v současné době jedním z neúčinnějších způsobů řešení složitých problémů [24].

## 4.2 Typy strojového učení s dohledem

Jelikož v této práci dále bude navrhován algoritmus, který řeší klasifikační problém, budou v této kapitole blíže popsány algoritmy, které se pro řešení tohoto typu problému používají nejčastěji.

### K-NN (K-Nearest Neighbors)

Algoritmus lze využít při řešení regrese i klasifikace. KNN je založený na výpočtu euklidovských vzdáleností v  $M$  rozměrném prostoru, kde hodnota  $M$  je dána počtem vstupních proměnných  $X$ . Při klasifikaci vzorku bez kategorie se výpočtem vzdáleností najde nejbližší vzorek z trénovacích dat, poté se neznámý vzorek zařadí do stejné kategorie. Tento základní postup se rozšiřuje o množství sousedů, podle kterých se určí výsledná kategorie. To znamená, že při určování kategorie algoritmus nevezme v potaz jenom nejbližšího souseda, ale  $N$  nejbližších sousedů a podle převažující kategorie mezi  $N$  ( $K$  v angličtině) sousedy určí kategorii neznámého vzorku.



Obr. 16 Jednoduché vysvětlení KNN algoritmu [25]

## LR (Logistic Regression)

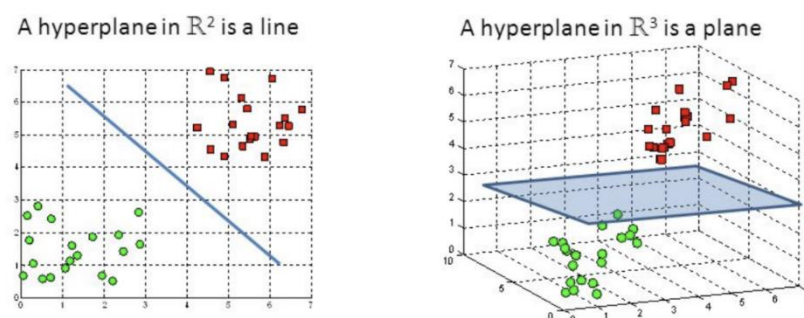
Logistická regrese je metoda zabývající se odhadem pravděpodobnosti jevu na základě vstupních dat vzorků. Pravděpodobnost jevu nabývá hodnot od 0, kdy jev nenastal, do 1, kdy jev nastal. Logistická regrese umožňuje určit proměnné, které ovlivňují, zda jev nastal nebo ne, a případně určit i míru jejich vlivu.

## LDA (Linear Discriminant Analysis)

LDA, kromě klasifikace a regrese, i redukcí dimenzí dat před dalším zpracováním. Cílem LDA je projekce vstupních dat do méně dimenzionálního prostoru se zachováním důležitých informací pro klasifikaci z důvodu vyšší odolnosti vůči přeučení modelu a snížení výpočetní náročnosti [26].

## SVM (Support Vector Machine)

Základním principem je najít rovinu v  $N$ -dimenzionálním prostoru ( $N$  – počet proměnných vzorku), který s jistotou rozdělí datové body do kategorií. K oddělení dvou kategorií datových bodů existuje mnoho možných rovin, které lze zvolit. Cílem je najít rovinu, která maximalizuje vzdálenost mezi datovými body obou kategorií. Maximalizace vzdálenosti zvyšuje pravděpodobnost správné klasifikace.



Obr. 17 Příklad rozdělovací roviny SVM ve dvou a tří rozměrném prostoru [27]

## RT (Random Tree)

RT je možné použít jak pro klasifikaci, tak i regresi. Algoritmus je hierarchický model se stromovou strukturou, která je tvořena dvěma typy uzlů, vnitřních (nelistových) a vnějších (listových). Každý z vnitřních uzlů obsahuje testovací funkci a podle jejího výsledku se vybere následující uzel, ve kterém se bude pokračovat. Tento proces začíná v kořenovém uzlu a

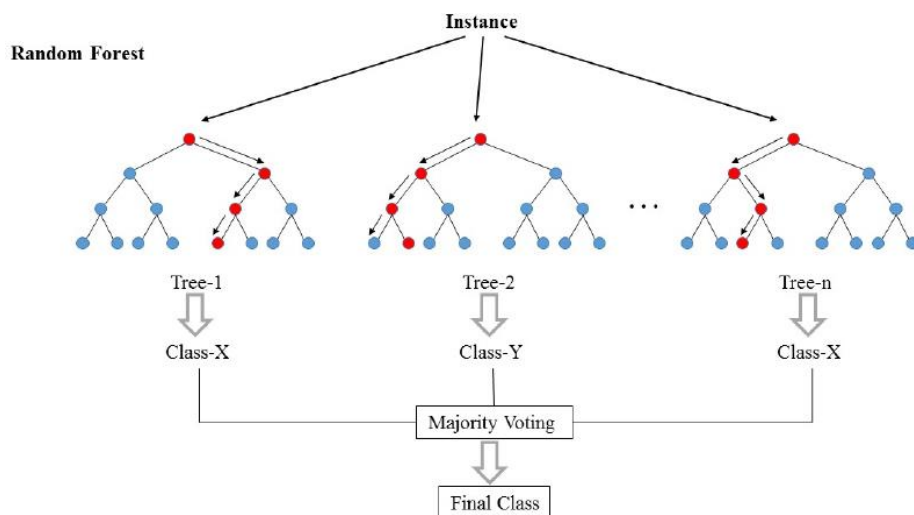
pokračuje rekurzivně až do dosažení vnějšího uzlu, který obsahuje číselnou hodnotu nebo označení třídy, do které bude zařazen vstupní vzorek.

Při tomto algoritmu je strom náhodným stromem vytvořeným náhodně ze všech možných stromů. Každý list obsahuje několik náhodných parametrů. Náhodné vytvoření stromu v tomto kontextu znamená, že každý strom v množině stromů má stejnou šanci výběru. Kombinace velkého počtu náhodných stromů obvykle vede ke správnému modelu.

## RF (Random Forest)

Algoritmus RF je založený na kombinaci několika rozhodovacích stromů. Každý strom závisí na hodnotách náhodného vektoru hodnot navzorkovaného nezávisle a se stejným rozložením pro všechny stromy v tzv. náhodném lese stromů.

Pro každý z rozhodovacích stromů se vytvoří náhodná sada trénovacích dat z originálních dat, k čemuž slouží funkce „bagging“. Pro každou sadu trénovacích dat je vytvořen strom, který je následně natrénován pouze na této sadě dat. Naopak při klasických stromech jsou uzly rozděleny na základě nejlepšího rozdělení mezi všemi proměnnými. Výsledná predikce je poté založena například na hlasování většiny mezi stromy v RF (tzv. Majority voting) při klasifikaci nebo průměrem hodnot při regresí.



Obr. 18 Schéma algoritmu RF [28]

### 4.3 Vyhodnocení modelu strojového učení

Jelikož modely strojového učení bývají typicky složitější než klasické modely, tzn. člověkem vytvořené modely, je pro vyhodnocení přesnosti modelu nutné výsledný model otestovat na testovacích datech, čímž se ověřuje úspěšnost modelu generalizovat vzory a vztahy mezi proměnnými. Výstupem hodnocení modelu určeného ke klasifikaci bývá typicky číselný odhad rozdílu v predikovaných a původních kategoriích vzorků, což bývá označováno jako tréninková chybovost. Dále model může být hodnocen podle toho, jestli je model přeučení nebo naopak nedoučený, podle míry schopnosti generalizovat vzory v trénovacích datech [29]. V této práci bude pro ověření modelu využita metoda křížové validace, pro úplnost budou v této kapitole popsány i další metody.

- A) Metoda zadržení (Holdout Method) – Datová sada je před trénováním modelu rozdělena na dvě části, trénovací a testovací sadu dat. Poměr mezi velikostmi sad dat je volitelný. Trénovací sada je použita k trénování modelu, testovací poté k jeho testování. Jedná se o jednoduchý typ křížové validační metody. Tato metoda trpí problémy s vysokou variabilitou výsledků z toho důvodu, že metoda nezohledňuje variabilitu vzorků v datové sadě. Výsledek může být při odlišném rozdělení dat na jednotlivé sady výrazně odlišný.
- B) K-násobná křížová validace (K-Fold Cross Validation) – Odstranění části dat pro ověření představuje možný problém s nedoučením. Redukcí tréninkových dat je riskována ztráta důležitých vzorů a vztahů v datové sadě, což zvyšuje možnost chyb. Tyto problémy řeší metoda K-násobné křížové validace, při které jsou data rozdělena do podmnožin  $k$  krát, kdy se jedna z podskupin  $k$  použije jako testovací sada a ostatní podmnožiny se spojí a vytvoří tréninkovou sadu. Každý datový vzorek se přesně jednou použije jako testovací a  $k-1$  krát jako trénovací. Úspěšnost modelu je získána zprůměrováním úspěšnosti v  $K$  iteracích, což významně snižuje zkreslení validace. Typicky se jako počet iterací používá  $K = 5$  nebo  $10$  [29].



Obr. 19 Schéma K-násobné křížové validace [30]

- C) Monte Carlo křížová validace (Monte Carlo Cross Validation) – Založena na stejném principu jako předchozí metoda, ale na rozdíl od ní, křížová validace Monte Carlo datový set rozděluje na trénovací a testovací sety náhodně, některé vzorky mohou být jako testovací vybrány několikrát, jiné na druhou stranu ani jednou. Počet iterací bývá násobně vyšší, typicky stovky [30].



Obr. 20 Schéma Monte Carlo křížové validace [30]

## 5. Rozbor RF lokalizačních systémů

V mnoha lokalizačních aplikacích uvnitř budov se prosadily systémy založené na Wi-Fi protokolu především díky rozšířenosti Wi-Fi infrastruktury. Obecně lze radiofrekvenční systémy rozdělit do dvou skupin podle jak přistupují k problematice určení polohy, a to na systémy založených na modelech šíření signálů nebo „otiscích“ (vzorcích) signálu.

### 5.1 Systémy založené na modelu šíření signálu

Modelové přístupy používají k určení polohy modely šíření rádiových vln prostorem. Ve venkovním prostředí je široce používán Friisův model šíření rádiových vln [31], který ale není vhodný v prostředích uvnitř budov z důvodu jejich složitosti a nehomogenního prostředí pro šíření RF signálu. Friisův vztah neboli rovnice ideálního přenosu je dána poměrem výkonů přijímače a vysílače.

$$\frac{P_p}{P_v} = G_v G_p \left( \frac{\lambda}{4\pi d} \right)^2$$

Rov. 1 Friisův vztah

Kde  $P$  je výkon vysílače/přijímače,  $G$  je zisk antény vysílače/přijímače,  $\lambda$  vlnová délka a  $d$  vzdálenost mezi vysílačem a přijímačem.

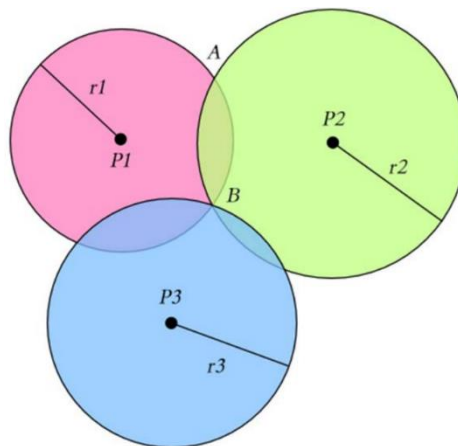
Při rozšíření Friisova modelu může být pro určení polohy v členitých vnitřních prostorech použita triangulace. Triangulace využívá měření vzdálenosti od referenčních bodů. Pro výpočet vzdálenosti mohou být použity různé metriky jako například Síla přijatého signálu (RSS), doba přijetí (ToA), rozdíl v čase přijetí (TDoA) nebo jejich kombinace [32].

Technology	Existing Hardware?	Min # anchors	Low Cost?	Res. (m) LOS	Res. (m) NLOS
RSSI	✓	3	—	1-3	5-10
TOA TDOA	✗	3	✗	0.2-0.4	1-5
AOA	✓	2	—	0.4	1-5
Time Reversal	✓	1	✓	0.02	0.02

Obr. 21 Porovnání technik pro výpočet vzdálenosti [33]

ToA a TR vyžadují časovou synchronizaci mezi zařízeními. Přesnost určení vzdáleností od referenčních zařízení je závislá na přesnosti měření času. Pokud se však mezi lokalizovaným zařízením a zařízeními na referenčních místech nachází překážky a signál se šíří vícecestní, přesnost lokalizace se výrazně snižuje [33].

Pro implementaci AoA, se typicky využívá anténní pole směrových antén, díky kterému je možné kromě síly signálu určit i směr, ze kterého se signál šíří. Výhody použití AoA jsou především ve venkovní lokalizaci, případně tam, kde je možnost komunikace na přímou viditelnost. Při použití ve vnitřních prostorách nebo v místě s vícecestným šířením lze algoritmus doplnit například výpočty sledování paprsku (Ray tracing) [34].



Obr. 22 Schéma triangulace [35]

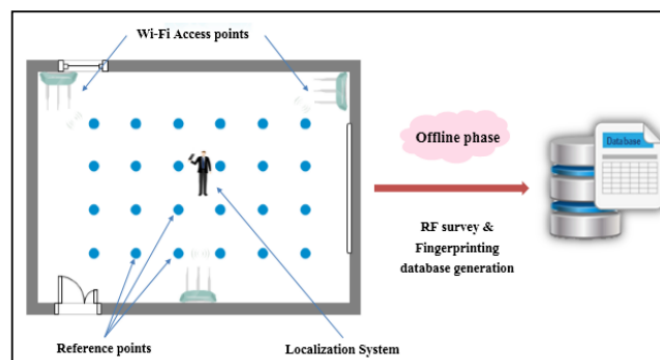
Na rozdíl od metody otisků, triangulace nepotřebuje velké množství referenčních dat v databázi a vystačí si s polohou referenčních bodů a jejich Mac adresami. Po výpočtu vzdálenosti od jednotlivých referenčních bodů v okolí je určena poloha podle Schéma triangulace na Obr. 22. Tato metoda určuje polohu uživatele v reálném čase a je tudíž náchylná k chybám při převodu parametrů signálu na vzdálenost nebo při jeho změnách způsobných průchodem neočekávanými překážkami. Výše uvedený model lze rozšířit modelem nelineární regrese a váhovacím algoritmem a zvýšit tím jeho přesnost a snížit chybovost [36]. Hlavním omezením těchto metod je, že jsou silně ovlivňovány nehomogenitami v prostředí a tím pádem i rozdílným šířením signálu od jednotlivých referenčních bodů. Přesnost lokalizace je dále ovlivněna vícecestným šířením s mnoha odrazy od překážek nebo pohybujících se osob.



## 5.2 Přístupy založené na otiscích signálu

Přístupy založené na otiscích signálu (Fingerprinting) mají typicky dvě fáze, a to off-line a on-line fázi.

1. V off-line fázi probíhá sběr dat, který je metodologicky i časově náročný z důvodu nutnosti sběru velkého objemu RSSI dat pro všechny referenční body. Nashromážděná data vytvoří RSSI mapu, s níž se poté v on-line fázi porovnávají naměřené hodnoty RSSI. Rozdíly hodnot RSSI pro jednotlivé referenční body nejsou závislé pouze na vzdálenosti od vysílače, ale také na překážkách ve směru šíření signálu. Změna topologie místa nebo pohyb osob může způsobit významné změny v rozložení RSSI. Tento problém lze kompenzovat například měřením prováděným za standardního provozu tzn. pohybu osob a předmětů v daném místě.

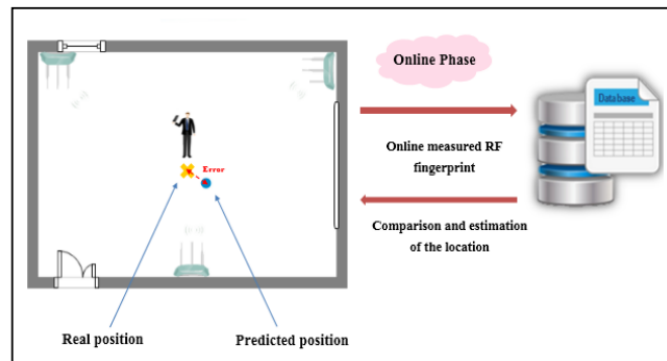


Obr. 23 Off-line fáze [37]

2. On-line fáze spočívá ve vyhledání referenčních bodů jejichž signálové otisky odpovídají nejvíce aktuálně měřeným hodnotám RSSI podle zvoleného algoritmu. Algoritmy lze rozdělit na deterministické a pravděpodobnostní.

Deterministické algoritmy využívají podobnosti signálu a naměřených otisků. Poloha je poté určena jako nejbližší referenční bod [38]. Pro srovnání signálu může být použita například euklidovská vzdálenost nebo kosinová podobnost. Hlavní výhodou deterministických metod je jejich snadná implementace a nízká výpočetní náročnost.

Pravděpodobnostní metody analyzují statistické vlastnosti databáze otisků a odvozují pravděpodobnost pozice měřeného RSSI. Ke zvýšení pravděpodobnosti nalezení správné pozice lze distribuce RSSI v referenčních bodech aproximovat parametricky, například s využitím Gaussova rozložení v určitých časových obdobích [39]. Pravděpodobnostní algoritmy bývají komplikovanější a vyžadují více referenčních dat než deterministické algoritmy.



Obr. 24 On-line fáze [37]

### 5.3 Vlastnosti lokalizačních systémů

Lokalizace ve vnitřních prostorech pokrývá rozmanité oblasti aplikací, pro které je nutné splnit různé požadavky. Mnoho vlastností systému může ovlivňovat parametry a relevantnost lokalizačních dat. Lze je vyjádřit například následovně: přesnost, odezva, pokrytí, adaptivita, škálovatelnost, cena a složitost [40].

- Přesnost – Hlavní kritérium pro posuzování lokalizačního systému. Může být definována jako velikost chyby mezi systémem odhadovanou polohou a skutečnou polohou uživatele, případně jako velikost nejistoty určení polohy. Čím je přesnost systému vyšší, tím je systém lepší.
- Odezva – Určuje, jak rychle může systém aktualizovat odhadovanou pozici pohybujícího se uživatele.
- Pokrytí – Pokrytí může být definováno jako lokální, škálovatelné a globální. Lokální pokrytí určuje jasně definovanou oblast, kterou nelze rozšířit (např. výrobní hala nebo budova). Škálovatelné pokrytí označuje schopnost rozšiřitelnosti pokrytí systému například přidáním hardwaru nebo zvýšením vysílacího výkonu. Globální pokrytí určuje přesnost systému mimo požadovanou oblast pokrytí nebo celosvětově.

- **Adaptivita** – Schopnost systému vypořádat se se změnami prostředí. Systém schopný kompenzovat změny může určovat polohu s vyšší přesností a předejít potřebě nové kalibrace systému.
- **Škálovatelnost** – Schopnost systému zvládnout různý objem požadavků a jeho potenciální rozšiřitelnost, v případě nárůstu požadavků na počet obslužených dotazů nebo pokrytí. V případě nárůstu požadavků může vést nízká škálovatelnost ke snížení výkonu systému, případně k nutnosti zásadních změn nebo duplikaci systému. Škálovatelný lokalizační systém by měl být schopen obslužit počet dotazů na polohu v dostatečně širokém intervalu bez dopadů na výkon.
- **Cena a složitost** – Náklady jsou definovány několika parametry, jako jsou náklady na instalaci nové potřebné infrastruktury, cena a údržba zařízení a náklady na elektrickou energii. Cenu zavedení může ovlivnit také nutnost kalibrace a její časová náročnost. Složitost systému se poté odvíjí od použitých algoritmů a výpočetních požadavků.

Zatímco systémy pro venkovní lokalizaci jsou široce rozšířené a využívané velkou částí populace, vnitřní lokalizace je v současné době stále ještě ve vývoji. V této části budou představeny různé technologie pro lokalizaci:

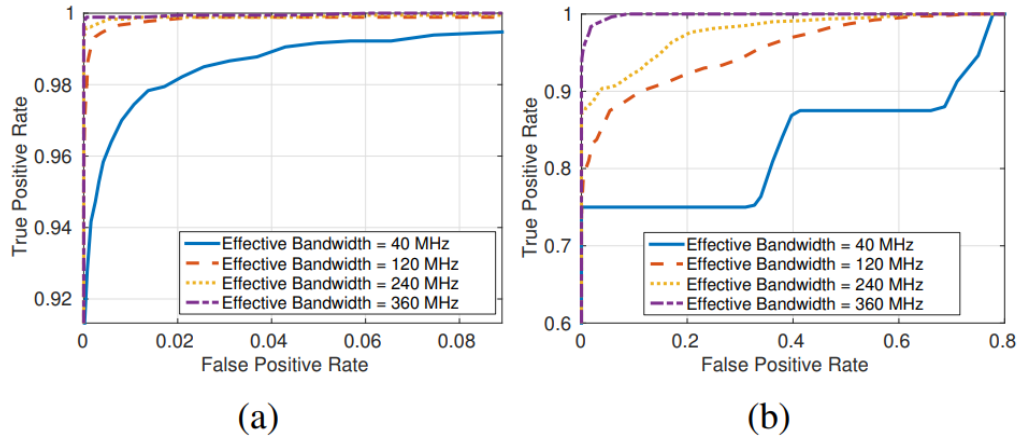
- **Globální navigační satelitní systém (GNSS)** je služba umožňující určování polohy za pomoci družic s celosvětovým pokrytím [41]. Uživatelé této služby používají malé elektronické rádiové přijímače, které na základě odeslaných signálů z družic umožňují vypočítat jejich polohu s přesností na desítky až jednotky metrů. Přesnost ve speciálních nebo vědeckých aplikacích může být až několik centimetrů. K získání pozice musí být přijat signál z minimálně čtyř satelitů.
- **Radar** vysílá opakující se signál pomocí anténního pole a pomocí druhého anténního pole přijímá signál odražený od sledovaných objektů, které musí být v přímé viditelnosti. Smíšením těchto dvou signálů vznikne mezi-frekvence, jejíž funkcí je vzdálenost. Díky dvěma odděleným anténním polím se dá sledovat poloha a rychlost více objektů najednou [42]. Do této kategorie se mohou řadit i systémy kombinující radiofrekvenční a ultrazvukovou technologii, které pro výpočet vzdálenosti připojených zařízení využívají časový rozdíl mezi příjmem RF a ultrazvukových signálů. Polohu je poté určena triangulací. Tyto systémy mají vysoké náklady, citlivost na hluk a nutnost komunikace na přímou viditelnost [43].

- Radiofrekvenční systémy určují lokaci z informace o síle přijímaného signálu (RSSI). Výpočet může využívat informace o RSSI z jednoho nebo více RF vysílačů umístěných tak, aby vytvořili překrývající se pokrytí v zájmové oblasti [44]. Počet potřebných vysílačů se odvíjí od aplikace a požadované přesnosti určení pozice. RF lokalizační systémy mohou využít stávající infrastrukturu, například Wi-Fi sítě, a nevyžadují přímou viditelnost mezi vysílačem a přijímačem. Existují metody, které umožňují kombinaci různých technologií pro zpřesnění určované polohy. Aplikace kombinující dva přístupové body Wi-Fi a dvě zařízení Bluetooth mohou dosáhnout přesnosti přibližně 1,5 metru [45].

Radiofrekvenční systémy jsou nejrozšířenější zejména kvůli nízké ceně a jednoduchosti aplikace. Mezi nejčastěji používané protokoly a oblasti jejich nasazení patří:

- RFID se prosadilo především ve výrobě a logistice. Díky bezkontaktnímu čtení zapsaných údajů je možné na definovaných přepravních trasách a linkách snadno sledovat pohyb výrobků. Dosah RFID pasivních tagů je v jednotkách cm, což v této aplikaci nepředstavuje omezení. V případě sledování pohybu po místnosti, mimo přepravní trasy, je tak nutné použít variantu aktivních RFID tagů, které mají dosah v desítkách metrů. V tomto případě zařízení komunikují na přímou viditelnost a lokalizace probíhá triangulací.
- Bluetooth se se zavedením nové verze radiového protokolu, Bluetooth Low Energy (BLE), prosazuje především v navigaci v budovách s využitím majáků a ultra-low power aplikacích, jako jsou například senzory. Bluetooth majáky mohou být umístěné na dveřích, stěnách, stropech a dalších místech, odkud vysílají rádiové signály v předem stanovených intervalech. Zařízení v dosahu pak mohou zachytit signál. Jeden maják v dosahu stačí například k určení patra nebo části budovy. Pokud jsou v dosahu přijímače alespoň dva majáky může být použit indikátor síly přijatého signálu (RSSI) pro lokalizaci pomocí fingerprintingu nebo k odhadu vzdálenosti mezi nimi pro triangulaci. Přesnost umístění se zvyšuje s počtem majáků.

- Wi-Fi má velkou výhodu v rozšířené infrastruktuře. Déle pak má větší dosah, než Bluetooth, typicky vyšší desítky metrů v závislosti na prostředí, a vyšší propustnost díky desítky megahertz širokému pásmu. S využitím fingerprinting metody v kombinaci s variabilní šířkou pásma, může být širší pásmo využito pro zvýšení rozlišení a úspěšnosti určování polohy [33].



Obr. 25 Závislosti úspěšnosti určení polohy na šířce pásma. a) s pohybem osob, b) s pohybem velkých předmětů

[33]

## 5.4 Současně oficiálně nabízené lokalizační služby v LPWAN sítích

Pro lokalizaci zařízení v aplikacích, kdy není k dispozici signál GNSS, v současné době Sigfox nabízí lokalizaci pomocí triangulace v síti Sigfox a službu Atlas, LoRa nabízí službu Collos.

Triangulace v síti Sigfox umožňuje určit pozici zařízení s přesností stovek až jednotek kilometrů, služba Atlas umožňuje lokalizaci na základě MAC adres Wi-Fi zařízení v okolí a zvýšit tak přesnost. Sigfox zařízení po zapnutí Wi-Fi modulu detekuje Wi-Fi zařízení v dosahu a jejich MAC adresy. Jelikož SigFox zpráva může pojmout maximálně 12 bajtů uživatelských dat musí se z MAC adres vybrat maximálně dvě adresy, které splňují dané parametry pro službu. MAC adresy jsou ve zprávě předány do Sigfox backendu, který je zpracuje pomocí služeb poskytovanými společnostmi Google a HERE Technologie. Lokalizace probíhá na základě porovnání přijatých MAC adres s databází s přístupovými Wi-Fi body a jejich souřadnicemi. Udávaná přesnost Sigfox Atlas je 500 metrů [46].

LoRa Collos je v době psaní této práce ve verzi preview a pracuje na podobném principu jako služba Atlas s tím rozdílem, že služba COLOS pro zpřesnění lokalizace kombinuje data o přístupových Wi-Fi bodech s RSSI, SNR a TOA posledních několika přijatých LoRa zpráv. Udávaná přesnost LoRa COLOS je 50 metrů [47].

## 6. Návrh a implementace algoritmu

---

Algoritmy pro lokalizaci v LPWAN sítích čelí různým výzvám. Mezi hlavní patří například limitace komunikační kapacity, vysílacího času, výpočetních možností a důraz na nízkou cenu senzorů. Z těchto důvodů navrhovaný algoritmus využívá pro lokalizaci RSSI fingerprinting Wi-Fi přístupových bodů v okolí. RSSI fingerprinting byl zvolen z cenových důvodů, jelikož oproti ostatním metodám nevyžaduje anténní pole nebo přesnou časovou synchronizaci. Technologie Wi-Fi byla zvolena z důvodu rozšířenosti infrastruktury a modulů Wi-Fi v IoT senzorech. To přináší možnost snazší implementace algoritmu do koncových zařízení a následně do zákaznických aplikací.

RSSI fingerprinting algoritmu se skládá ze dvou typických částí a to off-line a on-line fáze. Během off-line fáze se na referenčních místech naměří vzorky RSSI Wi-Fi přístupových bodů v komunikačním dosahu zařízení. Na základě těchto naměřených dat vytrénuje model strojového učení, podle kterého se bude poté v on-line fázi určovat poloha.

Jelikož se parametry jednotlivých LPWAN sítí užívaných pro IoT aplikace zásadně liší, musí být navrhovaný algoritmus přizpůsobitelný. Například z důvodu toho, že zařízení komunikující v síti SigFox nemohou odesílat ve zprávě víc než dvě MAC adresy, může být provedeno zpracování dat přímo v senzoru, čímž by se ušetřila energie potřebná k vysílání. Naopak zařízení komunikující v síti LoRa mohou mít méně výkonný hardware a naměřená data odeslat na zpracování do backendu.

### 6.1 Implementace

Navržený algoritmus byl implementován v jazyce Python verze 3.9.1 a pro přístup k datům ze síťové karty bylo využito Windows Wlan api rozhraní, jehož dokumentace je k dispozici na stránkách docs Microsoft [48].

Skrze Windows Wlan api rozhraní byla sbírána data o Wi-Fi přístupových bodech v okolí, a to konkrétně název sítě, MAC adresa, verze Wi-Fi protokolu, nosná frekvence, RSSI a LQI (Link Quality Indicator). Informaci o RSSI a LQI poskytuje ovladač Wi-Fi rozhraní pro určení kvality signálu. Hodnota RSSI je udávána v jednotkách decibelů vztažených na 1 miliwatt (dBm). Indikátor kvality spojení se pohybuje od 0 do 100, kdy hodnota 100 udává nejvyšší kvalitu spojení. Pro přepočítání mezi RSSI a LQI je dle dokumentace použita lineární interpolace [49]. Během testovacích měření byly zjištěny odchylky od tohoto vztahu. Z toho důvodu byly pro analýzy v kapitole 7 použity obě proměnné.

Pro algoritmy strojového učení byla využita knihovna *sklearn (scikit-learn)*, pro ukládání a načítání trénovaných modelů knihovna *pickle* a pro vytvoření časového plánu měřícího scriptu knihovny *sched* a *time*.

Program se skládá z více funkcí, jako hlavní funkce a jejich účel lze uvést:

- *wifi\_scan* – načtení vzorku dat o Wi-Fi přístupových bodech v dosahu
- *get\_sample\_....* – transformace dat z funkce *wifi\_scan*
- *get\_train\_data* – načtení naměřených dat
- *get\_model* – načtení trénovaného modelu strojového učení
- *(knn, rf, lda, svm) ...\_pipeline* – nastavení klasifikátoru modelu strojového učení
- *train\_model* – trénování modelu strojového učení
- *learn* – získání nového vzorku dat a jeho zápis
- *write\_data* – zápis naměřeného vzorku dat do souboru
- *locations* – výpis referenčních bodů a počet jejich naměřených vzorků
- *predict\_proba* – výpis pravděpodobnosti aktuální pozice podle naměřeného vzorku
- *crossval* – spuštění křížové validace náhodně seřazených naměřených dat

Naměřená data byla zapisována do textových souborů rozdělených do tří adresářů. Kromě adresářů pro soubory s hodnotami LQI a RSSI byl vytvořen adresář pro zálohování všech dat z funkce *wifi\_scan*. Pro každé referenční místo byl poté vytvořen jeden soubor v každém adresáři a jednotlivá měření byla zapisována do samostatných řádků ve formátu:

a) LQI:

```
{„SSID_1 MAC_1“: LQI_1, „SSID_2 MAC_2“: LQI_2, ..... „SSID_N MAC_N“: LQI_N}
```

b) RSSI:

```
{„SSID_1 MAC_1“: RSSI_1, „SSID_2 MAC_2“: ..... ,SSID_N MAC_N“: RSSI_N}
```

c) Záloha:

```
[ [„SSID_1“ „MAC_1“, „wifi version_1“, central_freq_1, RSSI_1, LQI_1], .....]
```

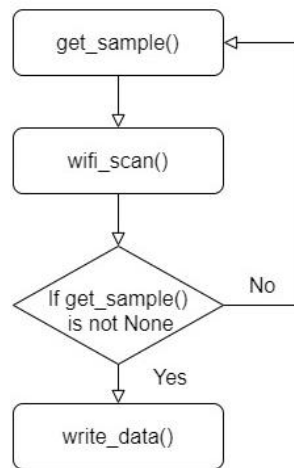
Tento formát byl zvolen pro svou přehlednost a vhodnost k dalšímu zpracování, kdy pro načtení jednotlivých řádků ze souborů s daty lze použít funkci *json.loads()*. Převod do formátu optimalizovaného pro trénování modelů strojového učení byla využita funkce *DictVectorizer()*. Vybrané části kódu jsou uvedeny v Příloha A.



## 6.2 Struktura algoritmu

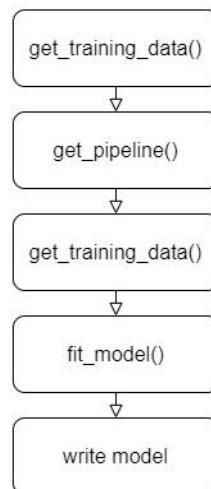
Program se dělí na dvě základní části. První část obsahuje měřící skript, který v definovaných časových intervalech spouští funkci *learn*. Tato funkce obsluhuje proces od měření dat až po jejich zápis do příslušných adresářů a souborů.

Funkce *wifi\_scan* s využitím Windows Wi-Fi api sbírá informace o Wi-Fi sítích v dosahu. Návratovou proměnou je pole obsahující *SSID*, *BSSID*, *type*, *CenterFreq*, *rsi*, *lqi*. Toto pole pak dále zpracovávají funkce *get\_sample\_...* (*lqi*, *rsi*, *all*), které data z pole převádí do formátu optimalizovaného pro další zpracování a zápis pomocí *write\_data*.



Obr. 26 Diagram funkce learn

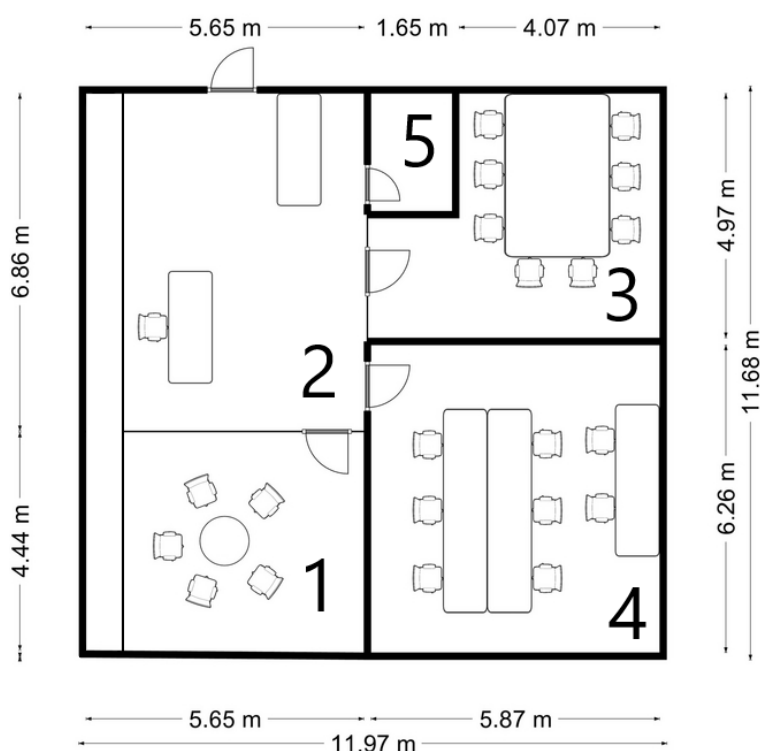
Funkce *train\_model* načte nastavení z funkce *...\_pipeline* a naměřená trénovací data. Pomocí funkce *fit\_model* spustí trénování modelu. Model se po trénování uloží do adresáře pomocí knihovny *pickle*.



Obr. 27 Diagram funkce train

## 7. Testování

Všechna měření v této kapitole proběhla na notebooku Dell XPS 15 9500 se síťovou kartou Killer Wi-Fi 6 AX500-DBS. Topologie testovací lokality je znázorněná na Obr. 28. Jednotlivé místnosti jsou označeny čísly 1 až 5, stěny mezi místnostmi číslo 1 a 2 a mezi 2 a 3 jsou tvořeny skleněnými příčkami. Testovací lokalita se nachází ve 23. patře kancelářské budovy City Tower na Pankráci v Praze. Při měření byla využita existující Wi-Fi infrastruktura bez znalosti umístění jednotlivých přístupových bodů nebo jejich aktuálního stavu. V dosahu testovacího zařízení bylo v dosahu typicky 5 až 15 Wi-Fi sítí.



Obr. 28 Topologie testovací lokality

Pro testování byly připraveny dva testovací scénáře, které budou více popsány v následujících kapitolách. Měření v obou testovacích scénářích probíhala na referenčních bodech v náhodném pořadí. V každém bodě proběhla dvě desetiminutová měření, která neprobíhala po sobě a při kterých byly vzorky ukládány každých 10 vteřin.

Na naměřených datech byly testovány 4 algoritmy, které jsou uvedeny níže. U jednotlivých algoritmů jsou uvedeny zkratky, které budou použité dále v testu, a jejich parametry. Pro přehlednost jsou uvedeny pouze parametry, které se liší od základního nastavení. Kompletní přehled parametrů nastavení je dostupný na stránkách dokumentace uvedených v referencích.

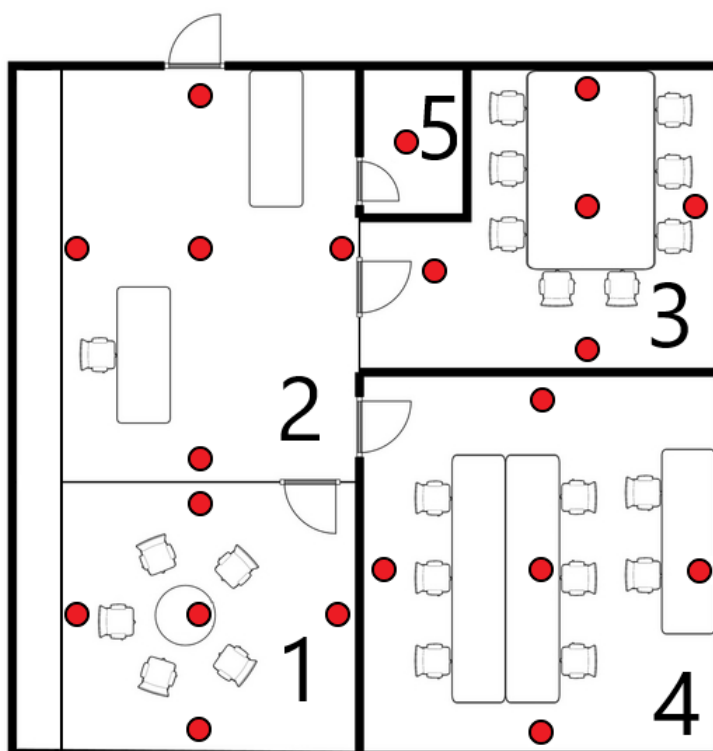
- RF – Random forest [50]  
Počet stromů v lese byl nastaven na 100 (`n_estimators = 100`) a váha jednotlivých kategorií na „vyváženo“ (`class_weight = "balanced"`).
- KNN – K-nearest neighbors [51]  
Počet sousedů, který vstupují do klasifikačního procesu byl nastaven na 10 (`n_neighbors = 10`).
- SVM – Support vector machine [52]  
Jako jádro algoritmu byla použita funkce radiálního základu (`kernel = "rbf"`). Byl zapnut výpočet pravděpodobnosti odhadu (`probability = "True"`).
- LDA – Linear Discriminant Analysis [53]  
Algoritmus LDA byl použit v základním nastavení.

## 7.1 Testovací scénář A

První testovací scénář měl určit vztah mezi počty referenčních bodů v místnosti a mírou jejího úspěšného určení pro testovací data. Měření referenčních bodů probíhalo půl metru od nejbližší zdi ve výšce jednoho metru. V místnostech číslo 1 až 4, bylo měřeno na celkem pěti referenčních bodech. Kvůli výrazně menším rozměrům místnosti č. 5 bylo v této místnosti měřeno pouze na jednom místě. Referenční body jsou na Obr. 29. označeny červenými body. Dále byla naměřena testovací sada šedesáti vzorků dat pro každou místnost v náhodných intervalech během nahodilého pohybu po místnosti. Tato testovací sada byla naměřena s týdenním odstupem. Před testováním nebyla tato data nijak upravována a případné změny v rozložení Wi-Fi infrastruktury v okolí nebyly nijak korigovány.

Pro další testování trénovaných modelů byla využita data měřená v rámci experimentu B, která tvoří naměřená data z 20 bodů v místnosti č. 2, celkem se jedná o 2400 vzorků.

Cílem testovacího scénáře bylo určit schopnost algoritmu správně určit místnost pro naměřené testovací vzorky.

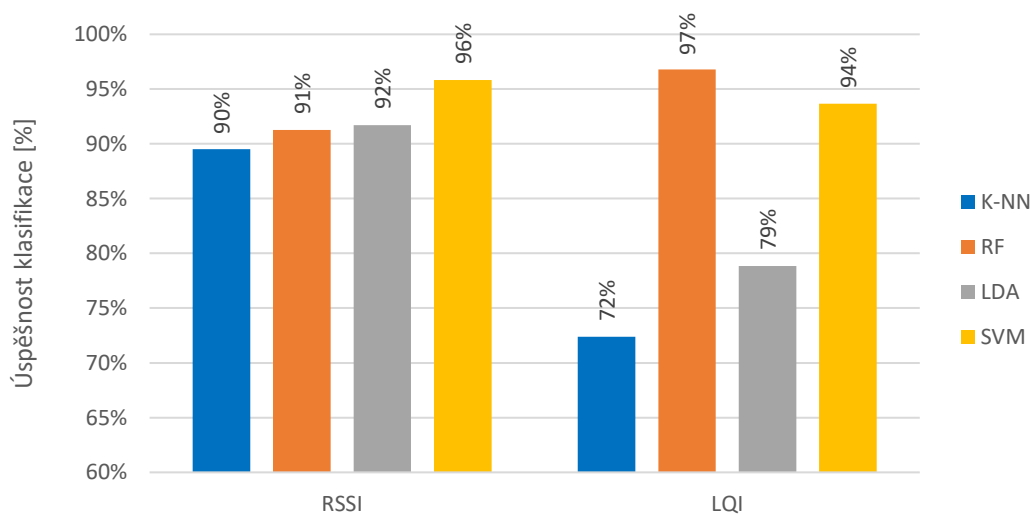


Obr. 29 Rozložení referenčních bodů pro jednotlivé místnosti

Z výsledků křížové validace naměřených dat v Tab. 1 je vidět, že kromě RF všechny testované algoritmy mají při použití LQI nižší úspěšnost. Dále je z výsledků křížové validace patrné, že SVM a RF vykazují při použití LQI i RSSI vyrovnanější výsledky a oproti zbylým dvěma algoritmům dosahují celkově vyšší úspěšnosti klasifikace.

run	K-NN		RF		LDA		SVM	
	RSSI	LQI	RSSI	LQI	RSSI	LQI	RSSI	LQI
1	0.890	0.768	0.909	0.968	0.914	0.792	0.954	0.936
2	0.894	0.731	0.918	0.970	0.913	0.788	0.959	0.937
3	0.898	0.721	0.909	0.968	0.922	0.787	0.959	0.937
4	0.895	0.720	0.911	0.964	0.915	0.785	0.960	0.937
5	0.898	0.725	0.917	0.970	0.921	0.790	0.959	0.936
<b>avg.</b>	<b>0.895</b>	<b>0.724</b>	<b>0.913</b>	<b>0.968</b>	<b>0.917</b>	<b>0.788</b>	<b>0.958</b>	<b>0.937</b>

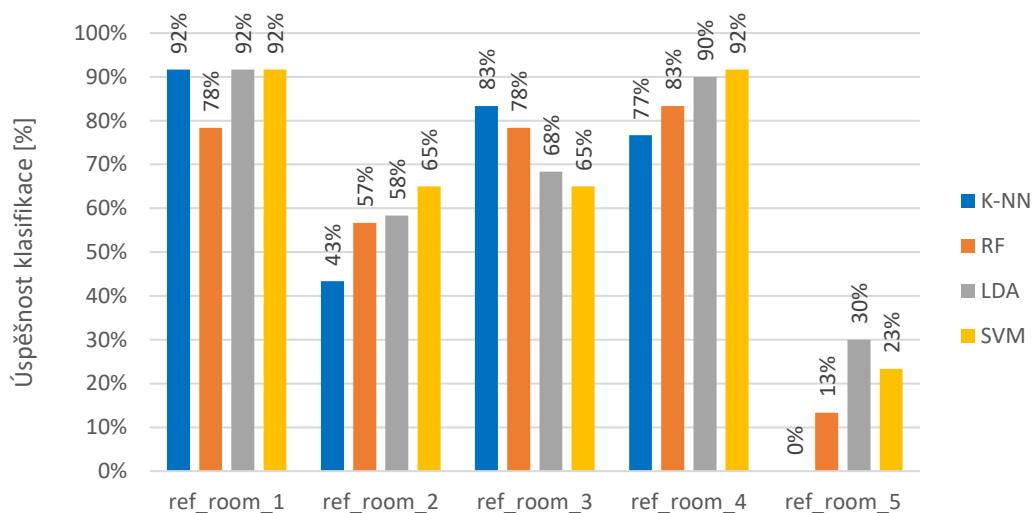
Tab. 1 Výsledky křížové validace pro trénovací data z referenčních bodů



Graf 1 Výsledky křížové validace pro trénovací data z referenčních bodů v procentech

	RSSI				
	ref_room_1	ref_room_2	ref_room_3	ref_room_4	ref_room_5
K-NN	55	26	50	46	0
RF	47	34	47	50	8
LDA	55	35	41	54	18
SVM	55	39	39	55	14

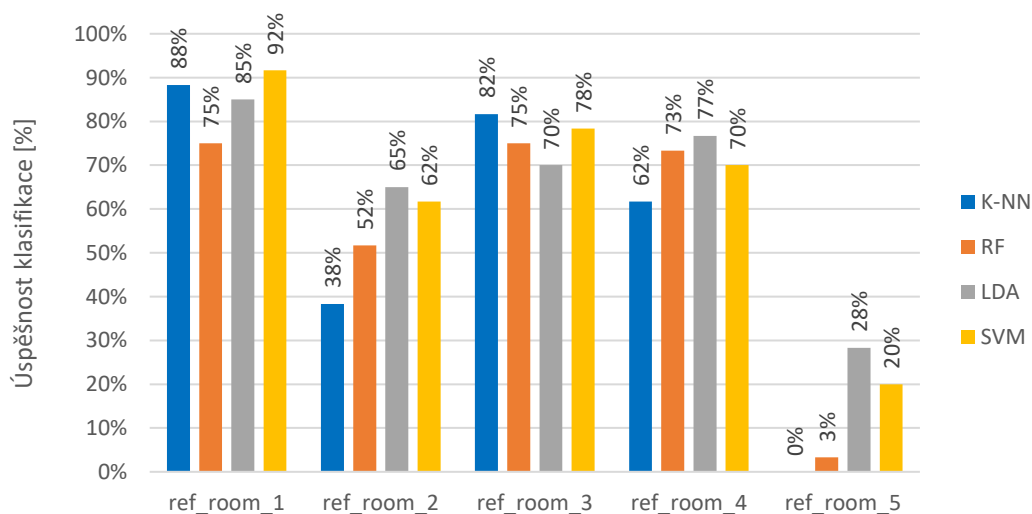
Tab. 2 Počet úspěšně klasifikovaných vzorků s využitím RSSI pro jednotlivé místnosti



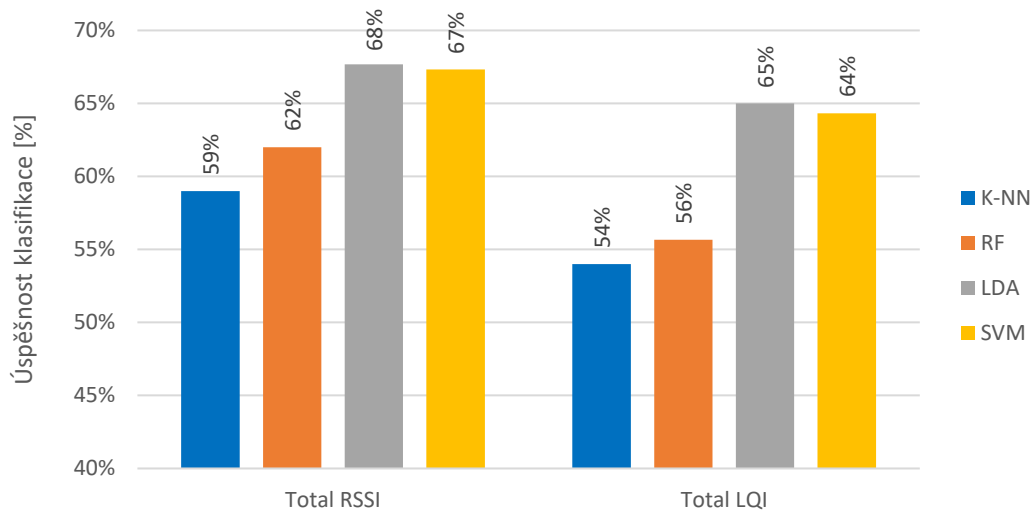
Graf 2 Úspěšnost klasifikace vzorků s využitím RSSI pro jednotlivé místnosti v procentech

	LQI				
	ref_room_1	ref_room_2	ref_room_3	ref_room_4	ref_room_5
K-NN	53	23	49	37	0
RF	45	31	45	44	2
LDA	51	39	42	46	17
SVM	55	37	47	42	12

Tab. 3 Počet úspěšně klasifikovaných vzorků s využitím LQI pro jednotlivé místnosti



Graf 3 Úspěšnost klasifikace vzorků s využitím LQI pro jednotlivé místnosti v procentech



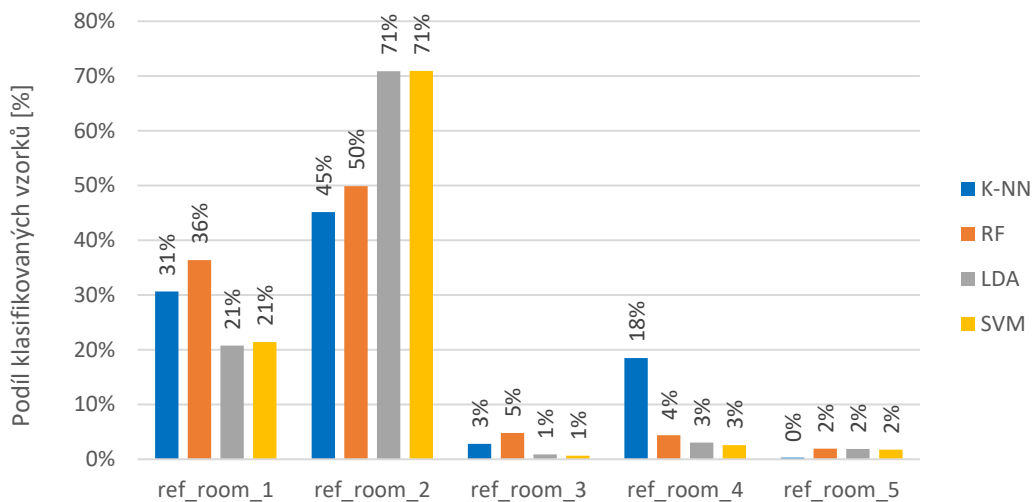
Graf 4 Průměrná úspěšnost klasifikace vzorků pro testovací sadu

Pro ověření přesnosti vytrénovaných modelů byla použita testovací sada 60 vzorků pro každou referenční místnost. Oproti křížové validaci je z výsledků vidět, že všechny testované modely na datech naměřenými s týdenním odstupem vykazují nižší úspěšnost klasifikace, a to jak při použití dat o LQI, tak RSSI. Při využití dat o RSSI vykazují všechny modely vyšší úspěšnost o 3 až 5 procent. Průměrnou úspěšnost klasifikace snižuje nejvýrazněji testovací sada pro místnost č. 5, pro kterou mají všechny testované modely úspěšnost v rozmezí 0 až 30 %, což může být zapříčiněno pouze jedním referenčním bodem v této místnosti, oproti ostatním místnostem, kde bylo naměřeno 5 referenčních bodů. Nejlepších výsledků s oběma testovacími metrikami dosahuje LDA společně s SVM jejichž úspěšnost se v průměru liší o jedno procento.

Pro další testování modelu pro určení místnosti byla využita naměřená sada dat pro experiment B, která obsahuje 2400 vzorků naměřených na 20 místech v referenční místnosti č. 2. V následujících tabulkách a grafech jsou uvedené výstupy jednotlivých klasifikačních algoritmů, jejich procentuální úspěšnost v určení referenční místnosti č. 2 pro všechny vzorky, a to jak v průměru, tak i v jednotlivých referenčních bodech v místnosti č. 2.

	RSSI				
	ref_room_1	ref_room_2	ref_room_3	ref_room_4	ref_room_5
K-NN	755	1112	70	455	8
RF	896	1229	119	108	48
LDA	511	1746	22	75	46
SVM	528	1748	16	64	44

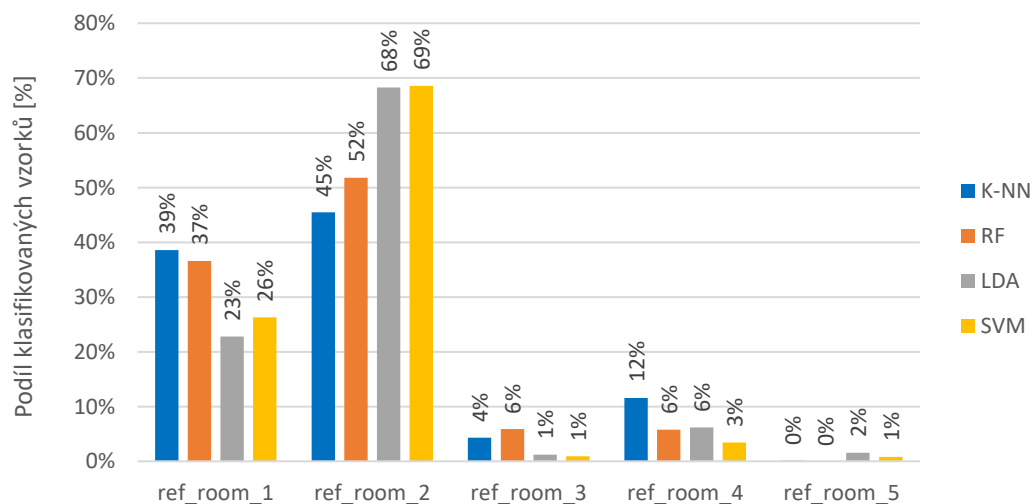
Tab. 4 Počet klasifikovaných vzorků jako jednotlivé místnosti s využitím LQI



Graf 5 Procentuální podíl vzorků klasifikovaných do jednotlivých místností s využitím LQI

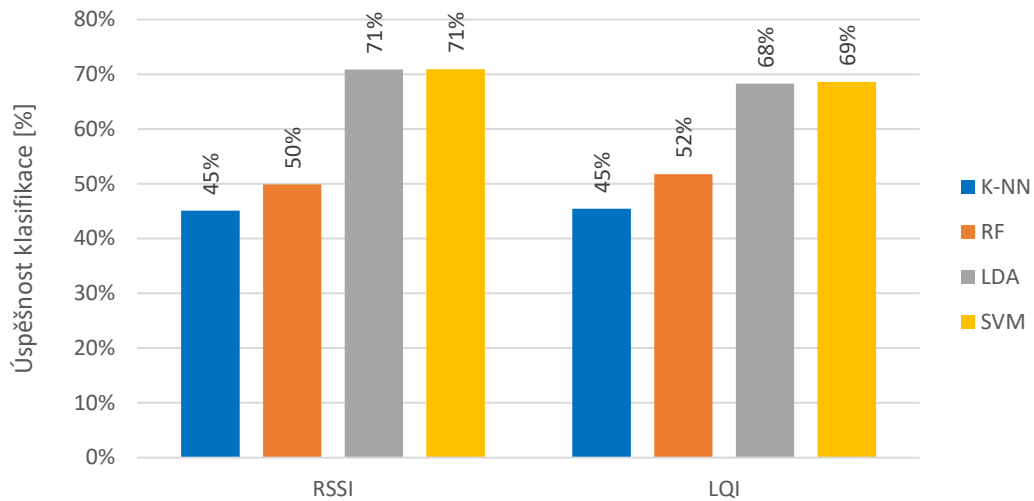
	LQI				
	ref_room_1	ref_room_2	ref_room_3	ref_room_4	ref_room_5
K-NN	926	1091	104	277	2
RF	878	1243	141	138	0
LDA	547	1639	29	148	37
SVM	631	1646	22	82	19

Tab. 5 Počet klasifikovaných vzorků jako jednotlivé místnosti s využitím LQI



Graf 6 Procentuální podíl vzorků klasifikovaných do jednotlivých místností s využitím LQI





Graf 7 Porovnání úspěšnosti klasifikace vzorků ze sady z experimentu B RSSI a LQI

Z grafu č. 7 vyplývá, že i u druhé sady testovacích dat mají LDA a SVM srovnatelnou úspěšnost klasifikace, která je zároveň zhruba o 20% vyšší než u zbylých dvou modelů. U druhé testovací sady se rozdíl mezi RSSI a LQI snížil na 1-3 %. Z tabulek č. 4 a 5 lze odečíst výsledky nesprávně klasifikovaných vzorků. Všechny modely nejvíce krát chybně zařadili vzorky do referenční místnosti č. 1, kterou od referenční místnosti č. 2 odděluje pouze skleněná příčka, kompletní rozložení testovacích místností je na Obr. 28, detail rozložení referenčních bodů pro experiment B v místnosti č. 2 je poté na Obr. 30.

Z Tab. 6, Tab. 7, Tab. 8 a Tab. 9 s úspěšností klasifikace testovaných modelů pro jednotlivé referenční body, nevyplývá, že by úspěšnost klasifikace byla závislá na poloze klasifikovaného bodu v rámci místnosti. Referenční místa s nejnižší průměrnou úspěšností klasifikace byla 2B, 4B a 5A pro RSSI a 4B, 4C a 5A pro LQI. Body 4B a 5A jsou společné pro obě metriky, což může indikovat náhodnou chybu při měření.

KNN/RSSI				
	A	B	C	D
1	48%	46%	58%	53%
2	32%	44%	52%	55%
3	45%	43%	54%	32%
4	46%	22%	35%	41%
5	34%	54%	59%	71%

KNN/LQI				
	A	B	C	D
1	46%	57%	50%	57%
2	35%	34%	57%	50%
3	54%	42%	43%	44%
4	43%	22%	36%	37%
5	27%	52%	58%	67%

Tab. 6 Úspěšnost klasifikace v jednotlivých referenčních bodech pro K-NN s využitím RSSI/LQI

RF/RSSI				
	A	B	C	D
1	69%	40%	44%	77%
2	56%	57%	59%	54%
3	61%	48%	81%	28%
4	48%	50%	22%	24%
5	51%	32%	40%	59%

RF/LQI				
	A	B	C	D
1	68%	57%	29%	85%
2	52%	32%	64%	38%
3	61%	72%	84%	59%
4	34%	37%	38%	28%
5	57%	38%	71%	75%

Tab. 7 Úspěšnost klasifikace v jednotlivých referenčních bodech pro RF s využitím RSSI/LQI

LDA/RSSI				
	A	B	C	D
1	80%	90%	66%	81%
2	89%	59%	67%	85%
3	77%	73%	92%	67%
4	74%	68%	57%	80%
5	50%	69%	71%	61%

LDA/LQI				
	A	B	C	D
1	72%	89%	62%	68%
2	79%	66%	77%	72%
3	78%	76%	83%	56%
4	72%	39%	56%	65%
5	43%	65%	70%	76%

Tab. 8 Úspěšnost klasifikace v jednotlivých referenčních bodech pro LDA s využitím RSSI/LQI

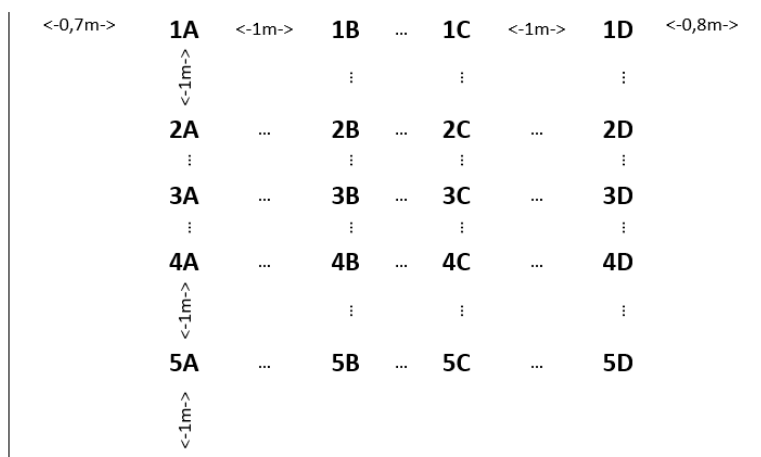
SVM/RSSI				
	A	B	C	D
1	73%	88%	74%	84%
2	72%	57%	68%	92%
3	78%	72%	86%	72%
4	68%	47%	66%	82%
5	52%	68%	73%	79%

SVM/LQI				
	A	B	C	D
1	60%	82%	50%	81%
2	63%	50%	73%	81%
3	77%	75%	78%	74%
4	68%	26%	60%	79%
5	55%	67%	70%	90%

Tab. 9 Úspěšnost klasifikace v jednotlivých referenčních bodech pro SVM s využitím RSSI/LQI

## 7.2 Testovací scénář B

Druhý testovací scénář měl za cíl určit teoretickou přesnost lokalizace v rámci jedné místnosti, celkem bylo naměřeno 20 referenčních bodů v rastru s hranou o velikosti 1 metr. Dále pak byl na všech 20 referenčních bodech, s odstupem dvou týdnů, naměřen testovací set dat o celkové velikosti 1200 vzorků, pro každé referenční místo 60 vzorků.

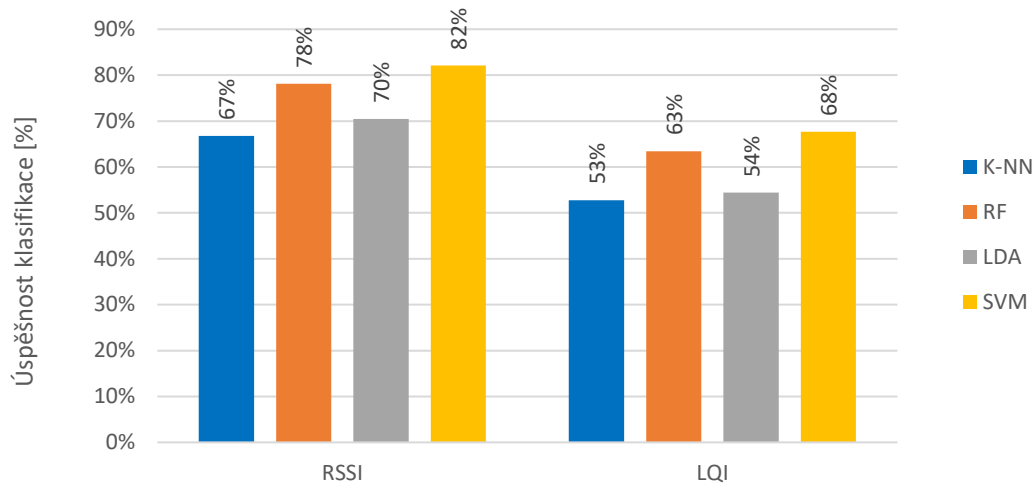


Obr. 30 Rozložení jednotlivých referenčních bodů v místnosti č. 2

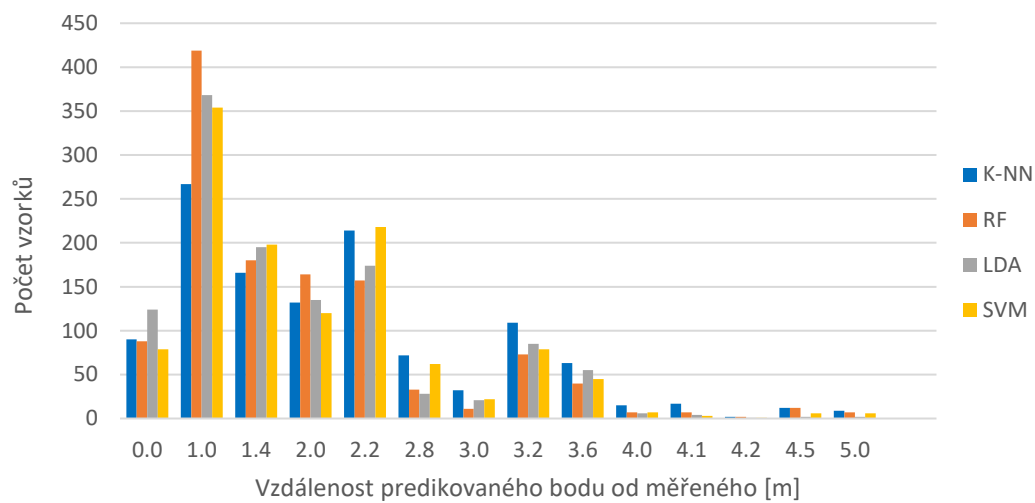
Křížová validace naměřených dat:

run	K-NN		RF		LDA		SVM	
	RSSI	LQI	RSSI	LQI	RSSI	LQI	RSSI	LQI
1	0.671	0.530	0.785	0.638	0.703	0.542	0.824	0.681
2	0.666	0.530	0.779	0.632	0.703	0.547	0.820	0.673
3	0.662	0.518	0.777	0.625	0.703	0.538	0.821	0.671
4	0.669	0.530	0.785	0.636	0.713	0.542	0.821	0.673
5	0.669	0.530	0.780	0.641	0.703	0.550	0.820	0.686
<b>avg.</b>	<b>0.667</b>	<b>0.528</b>	<b>0.781</b>	<b>0.634</b>	<b>0.705</b>	<b>0.544</b>	<b>0.821</b>	<b>0.677</b>

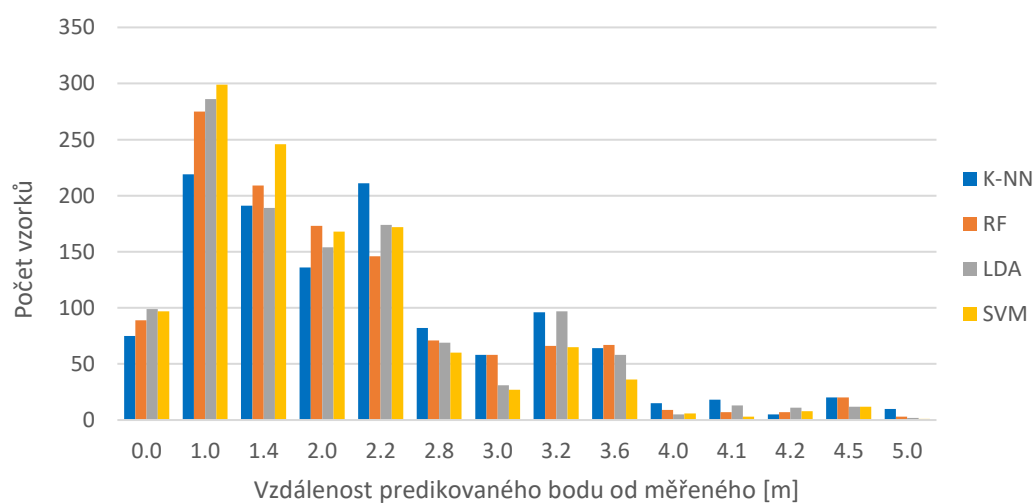
Tab. 10 Výsledky křížové validace pro data z referenčních bodů



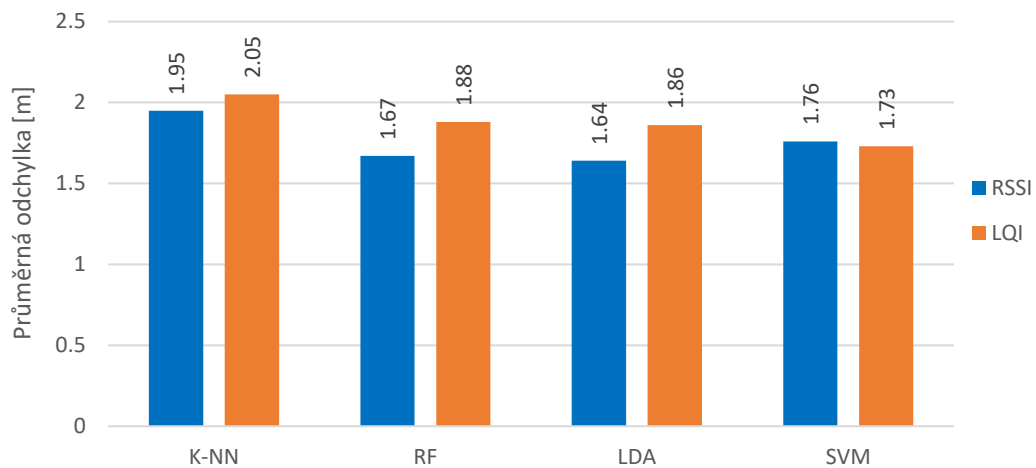
Graf 8 Výsledky křížové validace pro trénovací data v procentech



Graf 9 Distribuce odchylky od měřeného referenčního bodu při použití RSSI



Graf 10 Distribuce odchylky od měřeného referenčního bodu při použití LQI



Graf 11 Průměrná odchylka od měřeného ref. bodu

Z Graf 9, Graf 10 a Graf 11 vyplývá, že i v rámci experimentu B všechny testované modely, kromě SVM, dosáhly s využitím dat o RSSI vyšší přesnost než při využití LQI. Model využívající algoritmus K-NN predikoval polovinu testovacích vzorků s odchylkou do dvou metrů, zbylé modely poté s odchylkou do 1,4m. Celkově nejnižší odchylky predikovaných pozic od skutečných dosahovala LDA, která s testovací sadou dat s údaji o RSSI dosáhla průměrné odchylky 1,64m.

## 8. Závěr

---

Cílem diplomové práce bylo seznámit se s nejčastěji používanými metodami lokalizace LPWAN zařízení a vybrané metody implementovat a experimentálně ověřit. V rámci seznamování se s problematikou byl navržen algoritmus založený na metodě RSSI fingerprinting. Tato metoda se pro lokalizaci zařízení v LPWAN sítích, z metod v zadání, jeví jako nejvhodnější, jelikož zařízení již typicky mají potřebný hardware pro lokalizaci a nízká cena je u těchto zařízení klíčovým parametrem.

Navrhovaný algoritmus byl otestován dvěma experimenty. Během experimentů byly porovnány čtyři vybrané metody strojového učení se dvěma sadami dat, a to s daty o RSSI a LQI. Navrhovaný algoritmus pracuje se stejným typem dat jako aktuálně oficiálně nabízené lokalizační služby pro LPWAN jako je například Collos pro síť LoRa nebo Atlas pro síť Sigfox, tj. informace o Wi-Fi sítích v dosahu.

Měření byla prováděna s notebookem bez externí antény v prostředí bez znalosti rozmístění prvků Wi-Fi infrastruktury. Všechna testovací měření byla navíc prováděna s časovým odstupem, během kterého, v rozsáhlé kancelářské budově, pravděpodobně došlo k alespoň malým topologickým změnám, které ovlivňují šíření signálu v budově. I přes tyto limitace se v experimentu A, který měl ověřit schopnost algoritmu určit místnost, ve které se zařízení nachází, podařilo dosáhnout úspěšnosti klasifikace až 68 % pro první testovací sadu a 71 % pro druhou testovací sadu. Úspěšnost klasifikace pro první testovací sadu negativně ovlivnilo měření v místnosti č. 5, kde byl měřen pouze jeden referenční bod. Předpoklad, že vzhledem k výrazně menší velikosti oproti ostatním místnostem zde nebude nutné měřit na stejném počtu referenčních bodů, se nepotvrdil. Nejvyšší dosažená průměrná úspěšnost klasifikace by pro první testovací sadu bez výsledků z místnosti č. 5 byla přibližně o 10 % vyšší, tj. 78 %. Pro obě testovací sady v rámci experimentu A dosahovaly nejvyšší úspěšnosti metody LDA a SVM s daty o RSSI, rozdíl v jejich úspěšnosti byl přibližně 1 %. V experimentu B, který měřil průměrnou odchylku při určování pozice v rámci místnosti, se podařilo při využití LDA v kombinaci s daty o RSSI dosáhnout průměrné odchylky 1.64 m. Kombinace metody LDA s daty o RSSI byla při většině testů nejúspěšnější.

Navrhovaný algoritmus lokalizace pomocí Wi-Fi nevyžaduje GNSS modul, to umožňuje případně snížit cenu a velikost zařízení pro aplikace v rámci výrobních hal nebo kancelářských budov. Navíc oproti službám Collos a Atlas, které dosahují ve vnitřních prostorech přesnosti 50 až 200 m, nabízí výrazně vyšší přesnost lokalizace při použití stejného typu dat. Na diplomovou práci by bylo možné dále navázat otestováním implementace a vlastností navrženého algoritmu na reálném LPWAN zařízení.

Funkce *wifi\_scan*:

```
01: def wifi_scan():
*
*
26:     networks = customresize(avail_net_list.NetworkBSS,
                             avail_net_list.NumberOfItems)
27:
28:     for net in networks:
29:         lqi = (str(net.uLinkQuality))
30:         SSID = net.dot11Ssid.SSID[:net.dot11Ssid.SSIDLength]
                 .decode('utf-8')
31:         BSSID = ':'.join('%02x' % b for b in net.dot11Bssid).upper()
32:         rssi= str(net.lRssi)
33:         CenterFreq = net.ulChCenterFrequency/1000000
34:         type = net.dot11BssPhyType
35:
36:         if type == 7 : type = "802.11n"
37:         elif type == 8 : type = "802.11ac"
38:
39:         BSSI_Values.append([SSID, BSSID, type, CenterFreq, rssi, lqi])
40:
*
*
46:     return BSSI_Values
```



**Funkce *learn*:**

```

01: def learn(label):
02:     label_path1 = os.path.join(data_lqi, label + ".txt")
03:     label_path2 = os.path.join(data_rssi, label + ".txt")
04:     label_path3 = os.path.join(data_all, label + ".txt")
05:     try:
06:         new_lqi = get_sample_lqi()
07:         new_rssi = get_sample_rssi()
08:         new_all = get_sample_all()
09:         if new_all:
10:             write_data(label_path1, new_lqi)
11:             write_data(label_path2, new_rssi)
12:             write_data(label_path3, new_all)
13:     except:
14:         print("Something went wrong!")

```

**Funkce (*knn, rf, lda, svm*) ...*\_pipeline*:**

```

01: def knn_pipeline(estimators=10):
02:     knn = KNeighborsClassifier(n_neighbors=estimators)
03:     return make_pipeline(DictVectorizer(sparse=False), knn)

01: def rf_pipeline(estimators=100):
02:     rf = RandomForestClassifier(n_estimators=estimators,
03:                                class_weight="balanced")
04:     return make_pipeline(DictVectorizer(sparse=False), rf)

01: def lda_pipeline():
02:     lda = LinearDiscriminantAnalysis()
03:     return make_pipeline(DictVectorizer(sparse=False), lda)

01: def svm_pipeline():
02:     svm=SVC(kernel = 'rbf', probability=True)
03:     return make_pipeline(DictVectorizer(sparse=False), svm)

```

**Funkce *train\_model*:**

```

01: def train_model(path = data_path, model = model_file, pipeline):
02:     X, y = get_train_data(path)
04:     if len(X) == 0: raise ValueError("Can not find any data!")
06:     lp = pipeline
07:     lp.fit(X, y)
09:     with open(model, "wb") as f:
10:         pickle.dump(lp, f)

```

**Měřicí script:**

```

01: import sched, time
02: s = sched.scheduler(time.time, time.sleep)
03:
04: def do_measure(delay):
05:     print("Doing measurement...")
06:     s.enter(delay, 2, learn, argument=("reference_point_1",))
07:     s.run()
08:
09: for i in range(60):
10:     print("measuring time:", i+1)
11:     do_measure(10)

```

**Funkce *crossval*:**

```

01: def crossval(data = data_path, folds=10, n=5, clf=None, X=None, y=None):
02:     if X is None or y is None:
03:         X, y = get_train_data(data)
04:     if len(X) < folds: raise ValueError("Not enough samples!")
05:     tot = 0
06:     print("KFold folds={}, running {} times".format(folds, n))
07:     for i in range(n):
08:         shuffle_index = ShuffleSplit(n_splits=folds, test_size=0.3,
                                     random_state=i)
09:         res = cross_val_score(clf, X, y, cv=shuffle_index).mean()
10:         tot += res
11:         print("{} / {}: {}".format(i + 1, n, res))
12:     print("----- total -----")
13:     print(tot / n)

```

# Reference

---

- [1] G.M.Djuknic, „Geolocation and Assisted GPS,“ IEEE Computer, 2001.
- [2] S. Singh, „Internet of Things (IoT): Security challenges, business opportunities & reference architecture for E-commerce,“ IEEE, 2015.
- [3] D. Evans, „The Internet of Things How the Next Evolution of the Internet Is Changing Everything,“ Cisco, 2011.
- [4] IDC, „30 Billion Autonomous Devices By 2020,“ 2013.
- [5] Statista, „Statista,“ [Online]. [cit. 18.5.2021]. Dostupné z: <https://www.statista.com/statistics/880822/lpwan-ic-market-share-by-technology/>.
- [6] M. Frustaci a P. Pace, „Evaluating Critical Security Issues of the IoT World: Present and Future Challenges,“ IEEE, 2017.
- [7] G. Ferre a E. P. Simon, „Sigfox and LoRa PHY and MAC layers,“ University of Bordeaux, Bordeaux, 2018.
- [8] M. Saelens, J. Hoebeke, A. Shahid a E. De Poorter, Impact of EU duty cycle and transmission power limitations for sub-GHz LPWAN SRDs : an overview and future challenges, EURASIP JOURNAL ON WIRELESS COMMUNICATIONS AND NETWORKING, 2019.
- [9] F. Khan, A. ur Rehman a M. A. Jan, „Efficient Resource Allocation for Real time Traffic in Cognitive Radio Internet of Things,“ IEEE, 2019.
- [10] Sigfox, „SigFox,“ 2021. [Online]. [cit. 18.5.2021]. Dostupné z: <https://www.sigfox.com/en>.
- [11] Sigfox, „Sigfox Technical Overview,“ Sigfox, Labrège, 2017.
- [12] CRA, 2021. [Online]. [cit. 18.5.2021]. Dostupné z: <https://www.cra.cz/sluzby-iot>.
- [13] „The Things Network,“ [Online]. [cit. 18.5.2021]. Dostupné z: <https://www.thethingsnetwork.org/docs/lorawan>.
- [14] <https://www.electronics-notes.com/articles/connectivity/lora/what-is-lora-basics-m2m-iot.php>, „LoRa Physical Layer & RF Interface“.
- [15] K. Mekkia, E. Bajica a F. Chaxela, „A comparative study of LPWAN technologies for large-scale IoT deployment,“ ResearchGate, 2019.
- [16] J. Rejzek, „Lupa.cz,“ 2019. [Online]. [cit. 18.5.2021]. Dostupné z: <https://www.lupa.cz/clanky/wi-fi-6-prichazi-co-je-pod-poklickou-noveho-standardu/>.
- [17] J. Lloret, M. Taha a J. M. Jimenez, „Wireless Technologies for IoT in Smart Cities,“ ResearchGate, 2018.
- [18] electronics-notes, „Wi-Fi Channels, Frequencies, Bands & Bandwidths,“ 2020. [Online]. [cit. 18.5.2021]. Dostupné z: <https://www.electronics-notes.com/articles/connectivity/wifi-ieee-802-11/channels-frequencies-bands-bandwidth.php>.
- [19] K. Sjöberg, „Measuring and using the RSSI of IEEE 802.11p,“ Lund University, Measuring and using the RSSI of IEEE 802.11p.
- [20] P. Labs, „Pioneer Labs,“ [Online]. [cit. 18.5.2021]. Dostupné z: <https://pioneerlabs.io/insights/the-three-types-of-machine-learning-algorithms/>.

- [21] R. Gandhi, „Towards Data Science,“ 2018. [Online]. [cit. 18.5.2021]. Dostupné z: <https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a>.
- [22] S. Asiri, „Towards Data Science,“ 2018. [Online]. [cit. 18.5.2021]. Dostupné z: <https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623>.
- [23] S. Mishra, „Towards Data Science,“ 2017. [Online]. [cit. 18.5.2021]. Dostupné z: <https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a>.
- [24] B. Osiński a K. Budek, „deepsense.ai,“ 2018. [Online]. [cit. 18.5.2021]. Dostupné z: <https://deepsense.ai/what-is-reinforcement-learning-the-complete-guide/>.
- [25] T. Sewwandi, „Towards Data Science,“ 2020. [Online]. [cit. 18.5.2021]. Dostupné z: <https://towardsdatascience.com/predicting-cardiovascular-disease-using-k-nearest-neighbors-algorithm-614b0ecbf122>.
- [26] S. Raschka, „Sebastian Raschka,“ 2014. [Online]. [cit. 18.5.2021]. Dostupné z: [https://sebastianraschka.com/Articles/2014\\_python\\_lda.html](https://sebastianraschka.com/Articles/2014_python_lda.html).
- [27] R. Gandhi, „Towards Data Science,“ 2018. [Online]. [cit. 18.5.2021]. Dostupné z: <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>.
- [28] N. K. Ph.D., „LinkedIn,“ 2016. [Online]. [cit. 18.5.2021]. Dostupné z: <https://www.linkedin.com/pulse/random-forest-algorithm-interactive-discussion-niraj-kumar/>.
- [29] P. Gupta, „Towards Data Science,“ 2017. [Online]. [cit. 18.5.2021]. Dostupné z: <https://towardsdatascience.com/cross-validation-in-machine-learning-72924a69872f>.
- [30] R. Patro, „Towards Data Science,“ 2021. [Online]. [cit. 18.5.2021]. Dostupné z: <https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b>.
- [31] H. T. Friis, „A note on a simple transmission formula,“ IEEE, 1946.
- [32] M. Shchekotov, „Indoor Localization Method Based on Wi-Fi,“ ICWMC, 2014.
- [33] C. Chen, Y. Han, Y. Chen a K. J. R. Liu, „Indoor GPS with Centimeter Accuracy using WiFi,“ IEEE, 2016.
- [34] S. Wielandt, J.-P. Goemaere a L. D. Strycker, „Multipath-Assisted Angle of Arrival Indoor Positioning System in the 2.4 GHz and 5 GHz Band,“ IEEE, 2016.
- [35] M. E. Rusli, M. Ali a N. Jamil, „An Improved Indoor Positioning Algorithm Based on RSSI-Trilateration technique for Internet of Things (IOT),“ IEEE, 2016.
- [36] Z. Li, T. Braun a D. C. Dimitrova, „A Passive WiFi Source Localization System based on Fine-grained Power-based Trilateration,“ IEEE, 2017.
- [37] C. Houria, „Indoor Localization Using Wi-Fi Method Based on Fingerprinting Technique,“ IEEE, 2017.
- [38] S. He, „Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons,“ IEEE, 2016.
- [39] M. Youssef, A. Agrawala a A. U. Shankar, „WLAN location determination via clustering and probability distributions,“ IEEE, 2003.

- [40] Z. Farid, R. Nordin a M. Ismail, „Recent Advances in Wireless Indoor Localization Techniques,“ Selangor, 2013.
- [41] European GNSS agency, „GNSS market report, Issue 6,“ 2019.
- [42] A. Antonucci, M. Corra, A. Ferrari, D. Fontanelli, E. Fusari, D. Macii a L. Palopoli, „Performance Analysis of a 60-GHz Radar for Indoor Positioning and Tracking,“ IEEE, Trento, 2019.
- [43] G. Borriello, A. Liu, T. Offer, C. Palistrant a R. Sharp, „Wireless Acoustic Location with Room-Level Resolution using Ultrasound,“ ACM Press, 2005.
- [44] T.-Y. Tsai, C.-C. Hsu, H.-H. Chiang a W.-Y. Wang, „Mobile Localization-Based Service Based on RSSI Fingerprinting Method by BLE Technology,“ IEEE, 2018.
- [45] Y. Gwon, R. Jain a T. Kawahara, „Robust indoor location estimation of stationary and mobile users,“ IEEE, 2004.
- [46] Sigfox, „Sigfox,“ [Online]. [cit. 18.5.2021]. Dostupné z: <https://build.sigfox.com/geolocation-sigfox-atlas>.
- [47] Semtech, „LoRaCloud,“ 2020. [Online]. [cit. 18.5.2021]. Dostupné z: <https://www.loracloud.com/documentation/geolocation>.
- [48] Microsoft, „Microsoft,“ 2019. [Online]. [cit. 18.5.2021]. Dostupné z: <https://docs.microsoft.com/en-us/windows/win32/api/wlanapi/>.
- [49] Microsoft, „docs microsoft,“ [Online]. [cit. 18.5.2021]. Dostupné z: [https://docs.microsoft.com/en-us/windows/win32/api/wlanapi/ns-wlanapi-wlan\\_association\\_attributes](https://docs.microsoft.com/en-us/windows/win32/api/wlanapi/ns-wlanapi-wlan_association_attributes).
- [50] s.-l. developers, „sklearn - Random Forest Classifier,“ 2020. [Online]. [cit. 18.5.2021]. Dostupné z: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>.
- [51] s.-l. developers, „sklearn - KNeighbors Classifier,“ 2020. [Online]. [cit. 18.5.2021]. Dostupné z: <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>.
- [52] s.-l. developers, „sklearn SVC,“ 2020. [Online]. [cit. 18.5.2021]. Dostupné z: <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>.
- [53] s.-l. developers, „sklearn Linear Discriminant Analysis,“ 2020. [Online]. [cit. 18.5.2021]. Dostupné z: [https://scikit-learn.org/stable/modules/generated/sklearn.discriminant\\_analysis.LinearDiscriminantAnalysis.html](https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html).
- [54] F. Emmerling, J. M. Kim a M. Behmke, „Wi-Fi 6: Key Innovations and their Contributors,“ Juve Patent, 2020.
- [55] Primer, „Wi-Fi: Overview of the 802.11 Physical Layer and Transmitter Measurements,“ Tektronix.