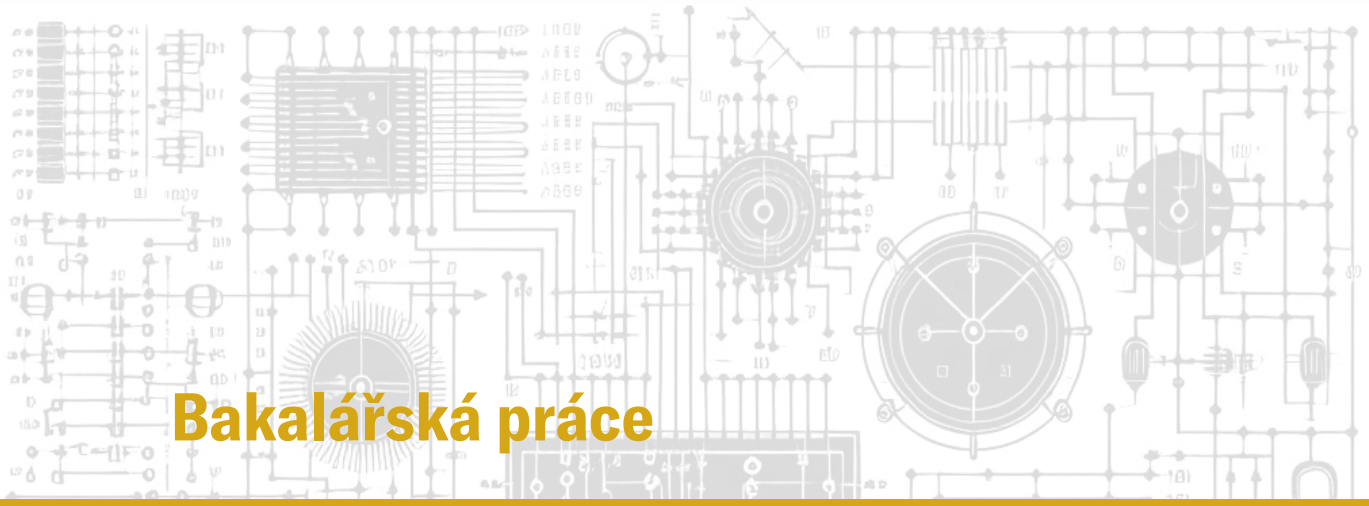
FAKULTA APLIKOVANÝCH VĚD
ZÁPADOČESKÉ UNIVERZITY
V PLZNI

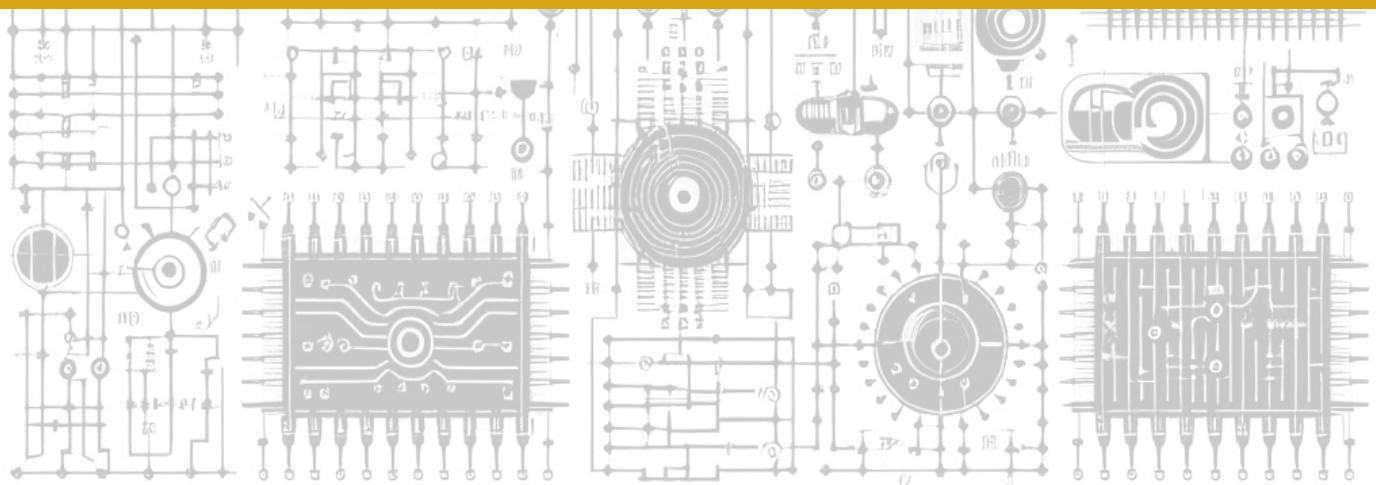
KATEDRA
KYBERNETIKY



Bakalářská práce

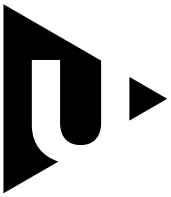
Predikce tranzitu elektrické energie z otevřených datových zdrojů

Martin Horešovský



PLZEŇ

2026



FAKULTA APLIKOVANÝCH VĚD
ZÁPADOČESKÉ UNIVERZITY
V PLZNI

KATEDRA
KYBERNETIKY

Bakalářská práce

Predikce tranzitu elektrické energie z otevřených datových zdrojů

Martin Horešovský

Vedoucí práce
Ing. Jan Švec, Ph.D.

PLZEŇ 2026

© Martin Horešovský, 2026

Všechna práva vyhrazena. Žádná část tohoto dokumentu nesmí být reprodukována ani rozšiřována jakoukoli formou, elektronicky či mechanicky, fotokopírováním, nahráváním nebo jiným způsobem, nebo uložena v systému pro ukládání a vyhledávání informací bez písemného souhlasu držitelů autorských práv.

Citace v seznamu literatury:
HOREŠOVSKÝ, Martin. *Predikce tranzitu elektrické energie z otevřených datových zdrojů*. Plzeň, 2026. Bakalářská práce. Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra kybernetiky. Vedoucí práce Ing. Jan Švec, Ph.D.

ZÁPADOČESKÁ UNIVERZITA V PLZNI
Fakulta aplikovaných věd
Akademický rok: 2025/2026

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení:	**Martin HOREŠOVSKÝ**
Osobní číslo:	**A23B0319P**
Studijní program:	**B0714A150005 Kybernetika a řídicí technika**
Specializace:	**Umělá inteligence a automatizace**
Téma práce:	**Predikce tranzitu elektrické energie z otevřených datových zdrojů**
Zadávající katedra:	**Katedra kybernetiky**

Zásady pro vypracování

1. Nastudujte problematiku modelování časových řad, včetně klasických metod a moderních přístupů založených na strojovém učení a neuronových sítích.
2. Proveďte rešerši dostupných otevřených datových zdrojů pro oblast energetiky a meteorologie (ENTSO-E Transparency Platform, ČEPS, OTE, CHMI a další).
3. Navrhněte a implementujte algoritmus pro zpracování a přípravu vstupních dat (časové sjednocení, imputace, tvorba příznaků).
4. Vyberte a natrénujte vhodné predikční modely pro tranzit elektrické energie přes hranice ČR a proveďte jejich vyhodnocení na základě vhodných metrik.
5. Zhodnoťte přínosy jednotlivých modelů, navrhněte možná vylepšení a diskutujte limity práce.

Rozsah bakalářské práce: **30-40 stran A4**
Rozsah grafických prací:
Forma zpracování bakalářské práce: **tištěná/elektronická**

Seznam doporučené literatury:

Dodá vedoucí práce.

Vedoucí bakalářské práce: **Ing. Jan Švec, Ph.D.**
Katedra kybernetiky

Konzultant bakalářské práce: **Ing. Petr Souček**
ČEPS, a.s.

Datum zadání bakalářské práce: **1. října 2025**
Termín odevzdání bakalářské práce: **25. května 2026**

Západočeská univerzita v Plzni
L.S.
Fakulta aplikovaných věd

doc. Ing. Miloš Železný, Ph.D.
děkan

doc. Dr. Ing. Vlasta Radová
vedoucí katedry

V Plzni dne 15. října 2025

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného akademického titulu.
K práci byly využity nástroje umělé inteligence Gemini CLI a Gemini na generování zdrojového kódu a úpravu, gramatickou kontrolu a formulaci textu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., zákon o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) v platném znění, a zejména skutečnost, že Západočeská univerzita v Plzni má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

V Plzni dne 25. května 2026

. . . . . .
Martin Horešovský

V textu jsou použity názvy produktů, technologií, služeb, aplikací, společností apod., které mohou být ochrannými známkami nebo registrovanými ochrannými známkami příslušných vlastníků.

Abstrakt

Dynamický rozvoj obnovitelných zdrojů energie v Evropě způsobuje rostoucí volatilitu v přenosových soustavách, což vede k častým a objemným neplánovaným tranzitním tokům elektrické energie přes území České republiky. Tato bakalářská práce se zabývá návrhem a implementací výpočetního systému pro predikci těchto fyzikálních toků s využitím metod strojového učení a otevřených datových zdrojů. Pro účely modelování byla plně automatizovaně získána a harmonizována data z platformy ENTSO-E Transparency a meteorologického API Open-Meteo. Problém byl koncipován jako regresní úloha nad časovými řadami.

V rámci experimentů bylo otestováno několik regresních modelů (například `ElasticNet`, `Lasso`, `SGDRegressor`) na odlišných sadách příznaků pro všechny čtyři přeshraniční profily ČR. Výsledky prokazují, že modely strojového učení dokážou velmi přesně aproximovat chování evropské sítě pouze na základě historických dat, a to i bez znalosti vnitřní fyzikální topologie. Nejlepších metrik bylo dosaženo při predikci celkové tranzitní bilance ČR s koeficientem determinace $R^2 = 0,9589$. Navržený systém byl rovněž úspěšně ověřen během několikatýdenního nasazení v reálném provozu. Práce ukazuje, že datově orientovaný přístup představuje efektivní nástroj pro krátkodobou predikci zatížení sítě, což může podpořit její bezpečnost i celkovou ekonomickou efektivitu dispečerského řízení.

Klíčová slova

predikce fyzikálních toků • přeshraniční toky elektřiny • strojové učení • časové řady • přenosová soustava • otevřená data • ENTSO-E • regresní analýza

Abstract

The dynamic expansion of renewable energy sources in Europe introduces increasing volatility into transmission grids, leading to frequent and massive unplanned electrical energy transit flows across the Czech Republic. This bachelor thesis focuses on the design and implementation of a computational system for predicting these physical flows using machine learning methods and open data sources. For modeling purposes, data from the ENTSO-E Transparency platform and the Open-Meteo API were automatically acquired and harmonized. The problem was framed as a time series regression task.

During the experiments, several regression models (e.g., `ElasticNet`, `Lasso`, `SGDRegressor`) were evaluated on various feature sets for all four cross-border profiles of the Czech Republic. The results demonstrate that machine learning models can highly accurately approximate the behavior of the European grid solely based on historical data, even without knowledge of the internal physical topology. The best performance was achieved in predicting the total transit balance of the Czech Republic, reaching a coefficient of determination $R^2 = 0.9589$. Furthermore, the proposed system was successfully validated during a multi-week deployment in a real-time operational environment. The thesis proves that a data-driven approach represents an effective tool for short-term prediction of grid load, which can enhance grid security and the overall economic efficiency of dispatch management.

Keywords

physical flow prediction • cross-border electricity flows • machine learning • time series • transmission system • open data • ENTSO-E • regression analysis

Poděkování

Na tomto místě bych rád poděkoval především svému vedoucímu práce, panu Ing. Janu Švecovi, Ph.D. a konzultantovi, panu Ing. Petru Součkovi. Velmi si cením všech předaných zkušeností, konzultací a poznámek během vypracovávání této bakalářské práce. Dále děkuji rodině za jejich podporu během celého studia a jejich trpělivosti. Další příbuzní a přátelé mi také byli oporou při studiu i v osobním životě. Děkuji celé katedře kybernetiky na Západočeské univerzitě za výuku během posledních tří let a mnoho předaných znalostí.

Obsah

1 Úvod a formulace problému **5**

1.1 Problematika tranzitních toků 5

1.2 Cíle práce 6

1.3 Výzkumné otázky 7

2 Teoretická část **9**

2.1 Energetická soustava 9

2.2 Datové zdroje 10

2.2.1 ENTSO-E Transparency platforma 10

2.2.2 Open-Meteo 10

2.3 Použité veličiny 10

2.3.1 Fyzikální toky (cíl predikce) 11

2.3.2 Charakteristika cílových dat 11

2.3.3 Plánované obchodní výměny 12

2.3.4 Výroba z obnovitelných zdrojů energie 12

2.3.5 Celková výroba 13

2.3.6 Skutečná a predikovaná zátěž 13

2.3.7 Ceny elektrické energie 13

2.3.8 Geografická specifika: Nabídkové zóny vs. Regulační oblasti 14

2.3.9 Meteorologická data (Open-Meteo) 14

2.3.10 Kalendářní a periodické příznaky 15

2.3.11 Sjednocení a časová rozlišení dat 17

2.4 Metody úpravy dat 18

2.4.1 Agregace dat 18

2.4.2 Výpočet salda 18

2.4.3 Rozdíl cen 18

2.4.4 Prostorové průměrování hodnot počasí 19

2.5 Základní vybrané metody modelování 19

2.5.1 `LinearRegression`, `Lasso`, `Ridge` 20

2.5.2 `LinearSVR` 20

2.5.3 `ElasticNet` . . . 21
2.5.4 `SGDRegressor` . . . 21
2.5.5 `HistGradientBoostingRegressor` . . . 21
2.6 Metody vyhodnocení . . . 21
2.6.1 Metriky vyhodnocování . . . 22

3 Návrh a implementace **25**
3.1 Implementace softwarového řešení . . . 25
3.2 Diagram projektu . . . 26
3.3 Načítání z datových zdrojů . . . 27
3.3.1 Dynamické načítání . . . 27
3.3.2 Ukládání trénovacích dat . . . 27
3.3.3 Využití ENUM konstant . . . 27
3.3.4 Komunikace s ENTSO-E API . . . 28
3.3.5 Komunikace s Open-Meteo API . . . 28
3.4 Příprava datasetu . . . 28
3.4.1 Pipeline proces . . . 28
3.4.2 Načítací skupiny . . . 28
3.4.3 Úpravy datasetu . . . 29
3.4.4 Další metody úprav . . . 29
3.5 Trénování modelů a vyhodnocení . . . 29
3.5.1 Vyhodnocovací metody . . . 30
3.5.2 Zobrazování výsledků . . . 30
3.6 Přechod na živé predikce . . . 31
3.6.1 Průběžné stahování hodnot z API . . . 31
3.6.2 Čas predikce . . . 31

4 Výsledky a vyhodnocení **33**
4.1 Výsledky pro jednotlivé hraniční profily . . . 34
4.1.1 Profil CZ → AT . . . 35
4.1.2 Profil CZ → DE_LU . . . 36
4.1.3 Profil CZ → PL . . . 37
4.1.4 Profil CZ → SK . . . 38
4.1.5 Celkový tranzit ČR . . . 38
4.1.6 Shrnutí výsledků . . . 39
4.2 Reálné predikce . . . 40
4.2.1 Metodika živého testování . . . 40
4.2.2 Kvantitativní výsledky . . . 40
4.2.3 Profil CZ → AT . . . 41
4.2.4 Profil CZ → DE_LU . . . 42

4.2.5 Profil CZ → PL 43
4.2.6 Profil CZ → SK 44
4.2.7 Celkový tranzit ČR 45
4.2.8 Shrnutí 46
4.3 Dopady kategorií příznaků na modely 47
4.3.1 Dominantní příznakové sady (C, N, E) 48
4.3.2 Význam autoregrese a počasí (A, B, W) 48
4.3.3 Negativní a zanedbatelný vliv (S, T) 49
4.3.4 Srovnání v grafu 50
4.4 Korelační analýza příznaků 52
4.4.1 Profil CZ → AT 53
4.4.2 Profil CZ → DE_LU 54
4.4.3 Profil CZ → PL 55
4.4.4 Profil CZ → SK 56
4.4.5 Shrnutí korelační analýzy 57

5 Závěr a shrnutí výsledků **59**
5.1 Shrnutí výsledků 59
5.2 Odpovědi na výzkumné otázky z Kapitoly 1.3 60
5.3 Zhodnocení predikce celkového tranzitu a budoucí rozvoj 61

Bibliografie **63**

Seznam obrázků **65**

Seznam tabulek **67**

Úvod a formulace problému

1

Současná energetika prochází dynamickou transformací směrem k decentralizaci a zvyšování podílu obnovitelných zdrojů energie (OZE), zejména větrných a fotovoltaických elektráren [Pet26]. Tento trend s sebou přináší zásadní změnu v dynamice výroby a přenosu elektrické energie. Na rozdíl od konvenčních zdrojů, které jsou řiditelné a jejichž výroba je plánovatelná s vysokou přesností, mají OZE stochastický charakter silně závislý na meteorologických podmínkách. Tato volatilita, v kombinaci s geografickým oddělením míst s největší potenciální výrobou (např. větrné parky v Severním moři) a center spotřeby (průmyslový jih Německa a střední Evropa), vytváří značný tlak na stabilitu a řízení přenosových soustav [Bol13].

1.1 Problematika tranzitních toků

Česká republika, jakožto klíčový tranzitní uzel v rámci propojené evropské soustavy, čelí v důsledku probíhajících změn častým a objemným přetokům elektrické energie. Podstata tranzitu spočívá v tom, že výkon soustavou pouze protéká, aniž by v ní byl vyroben či spotřebován. I přesto však přenosovou soustavu fyzicky zatěžuje, což zvyšuje riziko poruch a výpadků z důvodu vysokého vytížení prvků sítě.

Hlavním problémem je skutečnost, že elektrická energie se nešíří podle sjednaných obchodních kontraktů, ale řídí se Kirchhoffovými fyzikálními zákony. Tento fenomén vede ke vzniku toků, kdy elektřina obchodovaná uvnitř jedné zóny fyzicky protéká přes soustavy sousedních států [Hel18].

Provozovatelé přenosových soustav (TSO), v České republice společnost ČEPS, musí za těchto podmínek zajistit bezpečnost sítě a dodržování kritéria $N - 1$, které definuje bezpečnost sítě i v případě výpadku jakéhokoli jednoho prvku soustavy. Neplánované tranzitní toky mohou přetěžovat vedení, což vyžaduje nákladná nápravná opatření [Hel18].

Právě zde hraje klíčovou roli přesná predikce tranzitních toků, která přináší zlepšení v následujících oblastech:

- **Ekonomická efektivita:** Častý vysoký tranzitní tok způsobuje značné fyzikální ztráty v síti elektrické přenosové soustavy (EPS), které musí provozovatel uhradit. Přesná predikce tranzitu je proto klíčová pro efektivní nákup energie na pokrytí těchto ztrát, optimalizaci nákladů na redispečink a omezení nutnosti drahých investic do infrastruktury.

- **Bezpečnost soustavy:** Včasná identifikace rizik umožňuje předcházet kaskádovým poruchám a snižovat riziko výpadků způsobených vysokým zatížením.

- **Integrace trhu:** Spolehlivější odhady toků umožňují uvolnit vyšší kapacity pro přeshraniční obchod při zachování bezpečného provozu.

S rozvojem iniciativ v oblasti otevřených dat v posledních letech, jako platforma Evropské sítě provozovatelů přenosových soustav elektřiny (ENTSO-E) Transparency Platform [ENT26], se otevírá prostor pro využití metod strojového učení, které mohou tyto komplexní fyzikální vztahy aproximovat i bez detailní znalosti topologie celé evropské sítě.

1.2 Cíle práce

Hlavním cílem této bakalářské práce je návrh a implementace výpočetního systému pro predikci fyzických toků elektrické energie přes hranice České republiky s využitím volně dostupných datových zdrojů. Problém je koncipován jako regresní úloha v rámci modelování časových řad.

K dosažení hlavního cíle byly stanoveny následující dílčí kroky:

- Nastudovat problematiku modelování časových řad od statistických metod po hluboké učení.

- Provést rešerši otevřených datových zdrojů pro oblast energetiky (ENTSO-E, ČEPS, OTE) a meteorologie (Open-Meteo, ČHMÚ).

- Navrhnout a implementovat algoritmus pro předzpracování dat (čištění, synchronizace, imputace, tvorba příznaků).

- Vybrat, natrénovat a vyhodnotit vhodné predikční modely na základě relevantních metrik.

- Zhodnotit přínosy modelů a diskutovat limity dosažených výsledků.

1.3 Výzkumné otázky

Práce se snaží v rámci analýzy a experimentální části nalézt odpovědi na následující klíčové otázky:

1. **Jaké faktory nejvýznamněji ovlivňují tranzitní toky?** Analýza korelace mezi tranzitem a proměnnými, jako je výroba z OZE v Německu či ceny na denním trhu.
2. **Lze nahradit komplexní fyzikální modely datovými modely?** Ověření, zda metody strojového učení dokáží dostatečně přesně aproximovat fyzikální toky pouze na základě historických dat bez znalosti vnitřních parametrů sítě.
3. **Jaký vliv má zahrnutí počasí do modelu?** Zkoumání, zda zahrnutí meteorologických dat z konkrétních lokalit zlepší predikci oproti vynechání těchto dat.

Teoretická část 2

Kvalita a dostupnost dat jsou klíčovými faktory pro úspěšné modelování časových řad v oblasti energetiky. Pro účely této práce je nezbytné integrovat heterogenní datové sady zahrnující energetické parametry, meteorologické předpovědi a kalendářní informace. Tato kapitola podrobně popisuje jednotlivé zdroje, jejich technické parametry, časové rozlišení a relevanci pro predikci přeshraničních toků.

Zásadním aspektem celé analýzy je rozlišování mezi skutečně realizovanými hodnotami a hodnotami predikovanými. Zatímco skutečná data mohou být z důvodu kauzality použita pouze se zpožděním, predikované hodnoty představují jediné informace, které má model k dispozici v reálném čase při výpočtu předpovědi na budoucí období.

2.1 Energetická soustava

Evropská elektrizační soustava představuje jeden z největších propojených synchronních systémů na světě. Původně izolované národní sítě byly historicky postupně propojovány do regionálních bloků, aby se zvýšila spolehlivost dodávek a umožnila se vzájemná výpomoc v případě výpadku velkých výrobních bloků. V současnosti je provoz a rozvoj této propojené sítě koordinován organizací ENTSO-E. Základním pravidlem jakékoliv elektrizační soustavy je nutnost udržovat neustálou rovnováhu mezi výrobou a spotřebou v reálném čase, aby byla zachována nominální frekvence sítě (v Evropě 50 Hz).

Propojení národních soustav umožňuje propojení denních trhů s elektřinou. Elektřina jako komodita přirozeně proudí z oblastí s nižší cenou, často z oblastí s vysokým podílem výroby z OZE v danou chvíli, do oblastí s cenou vyšší. Fyzický tok elektřiny se však neřídí obchodními smlouvami, ale fyzikálními zákony. Vznikají tak rozdíly mezi naplánovaným obchodním tokem a skutečným fyzickým tokem. Pro země, jako je Česká republika, to znamená, že část kapacity přenosového vedení je spotřebována na transport tranzitní a exportní energie, což vyžaduje plánování a predikci ze strany dispečinku.

2.2 Datové zdroje

2.2.1 ENTSO-E Transparency platforma

Hlavním zdrojem dat z oblasti energetiky pro tuto práci je Platforma ENTSO-E Transparency [ENT26]. Tento webový portál byl zřízen na základě nařízení komise Evropské unie č. 543/2013 [Evr13], které nařizuje provozovatelům přenosových soustav transparentně publikovat klíčová data o trhu s elektřinou a fyzickém provozu sítí. Portál spravuje centrálně organizace ENTSO-E, přičemž jednotliví národní operátoři (TSO) a nominovaní organizátoři trhu s elektřinou (NEMO) jsou povinni do něj automatizovaně zasílat data. Platforma disponuje robustní dokumentací datových řad, které popisují metodiku měření a formáty. Poskytuje tak centralizovaný a standardizovaný přístup k množství otevřených informací.

2.2.2 Open-Meteo

Jelikož počasí silně koreluje s výrobou z OZE, jsou druhým nezbytným zdrojem meteorologická data. Pro účely této práce je využito volně dostupné webové API Open-Meteo poskytující historická i predikovaná data o počasí. Výhodou této služby je možnost okamžitého dotazování přímo ze skriptů strojového učení, bezplatný přístup v rámci nekomerčního využití a dostatečná geografická granularita. Zásadním faktorem při návrhu dotazování je zohlednění limitů počtu požadavků na rozhraní API, které vyžadují implementaci postupného stahování a ukládání do paměti pro již jednou načtené hodnoty [Zip23].

2.3 Použité veličiny

Z energetických dat jsou využity především řady reprezentující celkové toky na jednotlivých přeshraničních profilech. Tyto časové řady jsou cílem predikce, proto lze využít pouze jejich zpožděné hodnoty. Dále jsou využity časové řady popsané v dalších kapitolách. Důležitým faktorem pro vyhodnocování je úplnost dat. Přestože nařízení EU vyžaduje jednotnost, kvalita a dostupnost se u jednotlivých členských států mírně liší, což způsobuje výpadky v datových řadách vyžadující imputaci chybějících hodnot. Většina přenosových sítí se sice přesouvá na 15minutový interval řízení a obchodování, některé hodnoty jsou však stále dostupné pouze v hodinovém rozlišení.

Z meteorologických parametrů se vybírají ty s přímým dopadem na síť: rychlost větru (větrná produkce), sluneční svit (solární produkce) a teplota vzduchu (vliv na spotřebu kvůli vytápění či klimatizaci). Tyto faktory jsou klíčové pro zachycení fyzikálních příčin vzniku nadprodukce v Evropě a následných tranzitních přetoků.

2.3.1 Fyzikální toky (cíl predikce)

Fyzikální toky představují naměřené hodnoty předávaného činného výkonu přes jednotlivé hraniční profily v daném čase. Tyto hodnoty jsou dodávány přímo provozovateli přenosových soustav a představují cílovou proměnnou navrženého regresního modelu.

Z hlediska dostupnosti jsou tato data publikována s minimálním zpožděním, typicky do 15 až 60 minut po skončení měřeného intervalu. Je však nutné poznamenat, že sdružení ENTSO-E neposkytuje žádné oficiální predikce fyzikálních toků, což právě definuje prostor pro uplatnění modelů strojového učení v této práci.

Úlohu můžeme rozdělit na dva problémy. Predikce těchto fyzikálních toků na daném profilu (například CZ → AT), nebo celkového tranzitu. Celkový tranzit je pak definován jako suma všech jednotlivých profilů dané zkoumané země. Tato práce se zaměřovala na zkoumání vývoje pro Českou republiku.

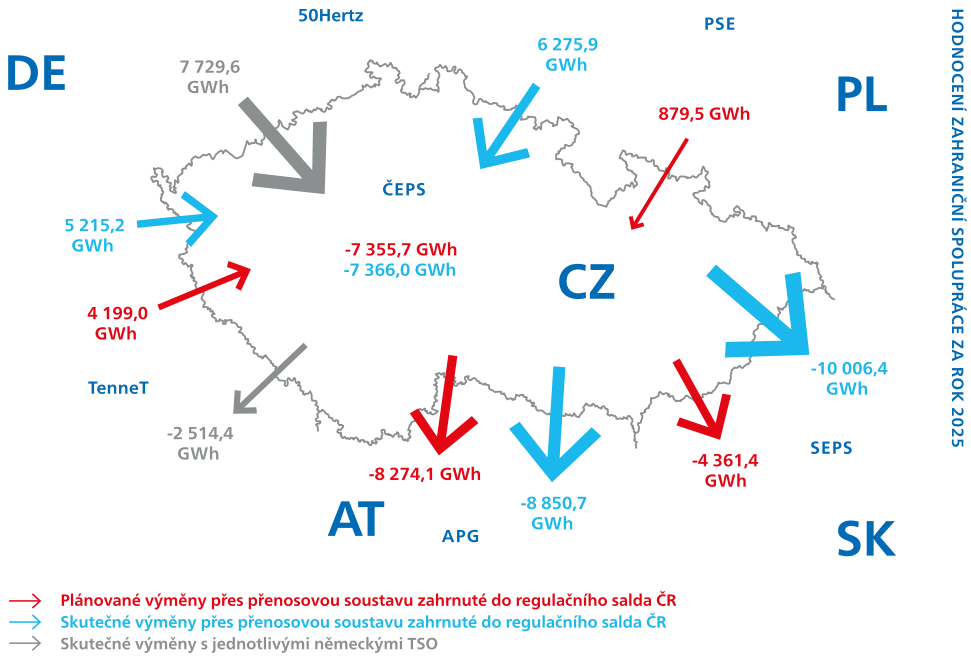
2.3.2 Charakteristika cílových dat

Pro lepší pochopení rozsahu predikované úlohy je nutné analyzovat fyzikální toky v kontextu reálných provozních dat. Na Obrázku 2.1 je znázorněna bilance zahraniční spolupráce přenosové soustavy ČR za rok 2025. Z hlediska predikce jsou klíčové rozdíly mezi plánovanými obchodními výměnami a skutečně realizovanými fyzikálními toky.

Z analýzy dat za rok 2025 vyplývají následující charakteristiky pro jednotlivé hraniční profily, které tvoří cílové proměnné modelu:

- **Profil CZ → AT:** Stabilní exportní profil s fyzikálním tokem 8 850 GWh, který je v relativně blízké shodě s plánovanými výměnami (8 274 GWh).
- **Profil CZ → DE:** Německé profily (50Hertz a TenneT) jsou charakteristické srovnatelnými toky v součtu, ale při jednotlivém pohledu jde vidět, že část energie ze severu na jih Německa protéká přes Českou republiku. Tyto objemy za rok 2025 činily 2 514 GWh.
- **Profil CZ → PL:** Výrazně tranzitní profil, kde fyzikální tok (6 275 GWh) převyšuje plánované hodnoty (pouhých 879 GWh) více než sedminásobně.
- **Profil CZ → SK:** Představuje významný exportní směr s celkovým fyzikálním tokem přes 10 000 GWh za rok. Rozdíl mezi plánem (cca 4,4 TWh) a skutečností (10 TWh) jasně ilustruje objem tranzitu směrem na jihovýchod a kruhové toky z Polska.

Ačkoliv výše uvedená mapa operuje s roční sumou energie (GWh), navržený model pracuje s okamžitými hodnotami činného výkonu v megawattech (MW).



Obrázek 2.1: Roční toky energie přes hranice ČR za rok 2025. Zdroj: [ČEP25]

2.3.3 Plánované obchodní výměny

Plánované obchodní výměny reprezentují agregované záměry účastníků trhu předávat energii mezi jednotlivými obchodními zónami. Tyto hodnoty odrážejí výsledky obchodování na denních a vnitrodenních trzích a slouží v modelu jako příznak, neboť vyjadřují záměr účastníků přesunout energii mezi zónami, přičemž tyto hodnoty korelují s budoucím fyzickým zatížením sítě.

Plány denní výměny vznikají na základě výsledků denních aukcí a jsou zveřejňovány obvykle v 13:00 SEČ pro následující den. Signalizují základní očekávanou bilanci mezi exportními a importními oblastmi.

2.3.4 Výroba z obnovitelných zdrojů energie

Vzhledem k rostoucímu podílu větrných a solárních elektráren v Evropě, a zejména v sousedním Německu, je jejich výroba kritickým faktorem ovlivňujícím tranzitní toky. Data jsou na platformě dostupná ve třech kategoriích:

1. **Predikce na zítřek:** Zveřejňována do 18:00 předchozího dne. Tato hodnota je pro model klíčovým vstupem, protože určuje očekávanou výrobu energie v systému v závislosti na meteorologických podmínkách.

2. **Vnitrodenní predikce:** Aktualizované odhady většinou ráno daný den, které reagují na zpřesněné meteorologické modely v čase blízkém reálnému provozu.

3. **Skutečná realizovaná výroba:** Energie skutečně vyrobená obnovitelnými zdroji není přímo vhodná pro trénování modelů. V čase predikce není tato informace dostupná, model by se tak naučil predikovat hodnotu na základě informace, která ještě nebyla naměřena.

2.3.5 Celková výroba

Dostupná jsou nejen data pro výrobu z obnovitelných zdrojů, ale i pro celkovou výrobu napříč energetickou sítí. Tato predikce je poskytována obvykle v 18:00 SEČ. Průběžné hodnoty skutečné výroby jsou aktualizovány periodicky v rámci obchodních intervalů.

2.3.6 Skutečná a predikovaná zátěž

Spotřeba představuje druhou stranu bilanční rovnice soustavy. Model pracuje s oficiální predikcí celkového zatížení pro každou obchodní zónu, která bývá k dispozici již kolem 10:00 předcházejícího dne. Odchylka mezi touto predikcí a skutečným naměřeným zatížením často vyvolává neplánované fyzikální toky, které musí přenosová soustava vyrovnávat.

2.3.7 Ceny elektrické energie

Ceny elektrické energie na denním trhu představují klíčový ekonomický indikátor, který motivuje pohyb energie v rámci evropské sítě. V systému propojených trhů dochází k automatickému přidělování přeshraničních kapacit tak, aby elektřina proudila z oblastí s nižší cenou do oblastí s cenou vyšší.

Data jsou na platformě ENTSO-E dostupná v rozlišení dané obchodní zóny a jsou zveřejňována obvykle krátce po 13:00 SEČ pro následující den. Přestože ceny přímo neurčují fyzikální zákony šíření energie, poskytují modelu informaci o ekonomických tlacích, které kaskádovitě ovlivňují zatížení přenosových tras.

2.3.8 Geografická specifika: Nabídkové zóny vs. Regulační oblasti

Zásadním aspektem pro korektní interpretaci dat je rozlišení mezi tržním a fyzikálním pohledem na síť. Nabídková zóna je oblast s jednotnou cenou, zatímco regulační oblast spravuje konkrétní provozovatel (TSO). Tento rozdíl je viditelný u Německa, které tvoří jednu obchodní zónu DE-LU, ale fyzicky se skládá ze čtyř regulačních oblastí (50Hertz, Amprion, TenneT, TransnetBW). Při modelování toků na hranici ČR–Německo je tedy nutné agregovat fyzikální toky na profilech ČEPS–50Hertz a ČEPS–TenneT, zatímco obchodní data jsou uvažována pro celou relaci CZ–DE. Používaná programová knihovna ovšem umožňuje tyto profily agregovat automaticky [Ene26].

2.3.9 Meteorologická data (Open-Meteo)

V práci je využito pokročilé rozhraní Open-Meteo, které agreguje data z globálních a regionálních numerických modelů. Tento zdroj poskytuje data s hodinovým rozlišením, která jsou následně lineárně interpolována na čtvrthodinové kroky.

- **Model ERA5 (Reanalýza):** Tato data představují historicky nejvěrnější rekonstrukci počasí na základě měření a zpětných modelových výpočtů. V práci ovšem využita nejsou. Jejich využití při trénování by nejspíše zpřesnilo model, ale ten by nezachycoval chybu predikce počasí. Vznikal by tedy větší nežádoucí rozdíl mezi metrikami při trénování a reálném použití.

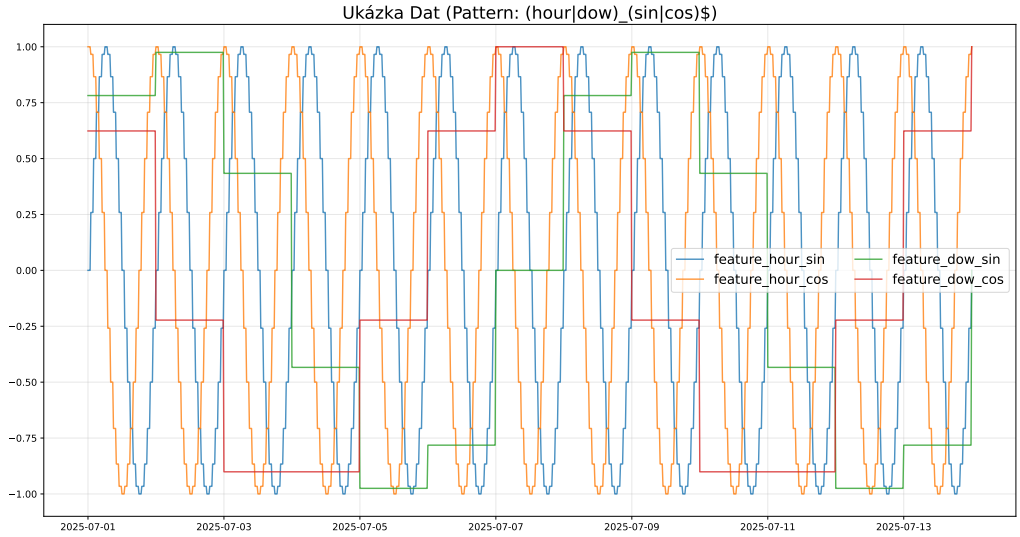
- **Historické předpovědi (Historical Forecasts):** Jsou zásadní pro realistické trénování modelů strojového učení. Na rozdíl od reanalýz obsahují tyto řady chyby a nejistoty přítomné v čase predikce, které jsou typické pro reálné předpovědi v čase D-1, tedy v době predikování den před predikovaným časem. Trénování na těchto datech učí model robustnosti vůči nepřesnostem předpovědí.

- **Předpovědi počasí:** Jedná se o reálně používané časové řady, které jsou dostupné v čase predikce. Open-Meteo poskytuje předpověď až 14 dní do budoucnosti, pro predikci jsou však využity pouze předpovědi na následující den, které jsou vhodnější a přesnější.

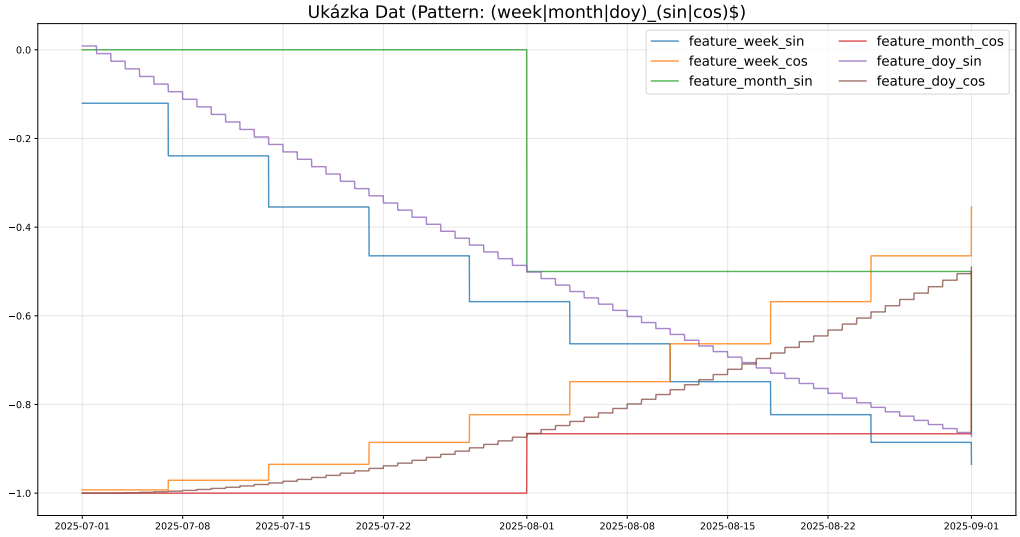
- **Sledované veličiny:** Klíčovými veličinami jsou teplota ve 2 metrech nad zemí (vliv na spotřebu), rychlost větru v 80 metrech (odpovídá výrobě větrných turbín), intenzita slunečního záření (vliv na výrobu solárních elektráren) a vlhkost a srážky pro zachycení obecných jevů počasí a rozšíření datové sady.

2.3.10 Kalendářní a periodické příznaky

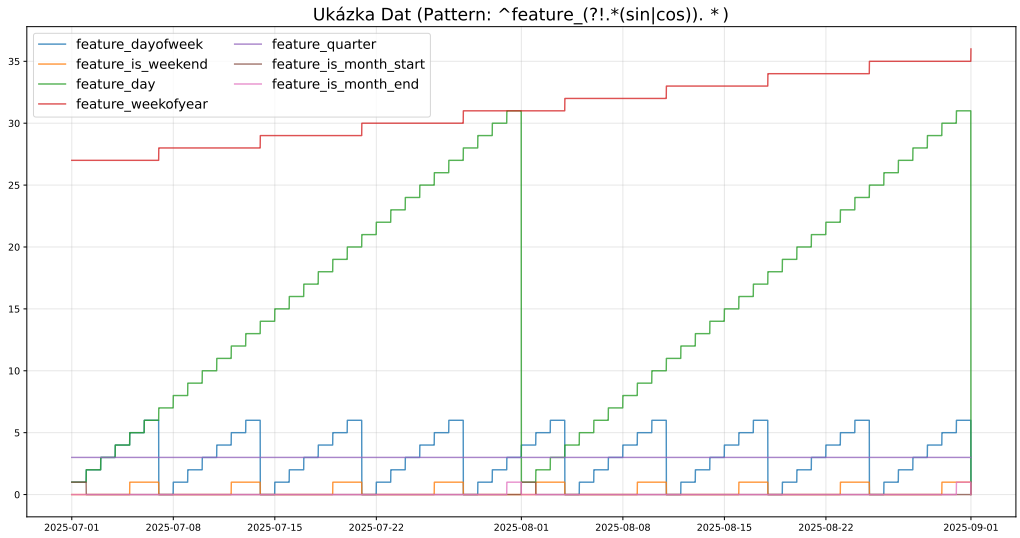
Chování elektrické sítě vykazuje silnou periodicitu závislou na lidské činnosti. Do modelu jsou proto zahrnuty odvozené příznaky, které tyto vzorce zachycují bez nutnosti jejich učení z čistých časových řad. Jedná se o různé veličiny rozsahem od hodiny až po měsíce či roky. Tyto hodnoty nejsou modelu předkládány pouze přímo, ale i v podobě sinové a kosinové transformace. Tyto transformace aplikují funkci sinus a cosinus na danou časovou řadu. Hodnoty se tím normalizují na interval od mínus jedné do jedné. Model tak dokáže lépe periodicitu zpracovat (Obrázky 2.2 a 2.3). Dále jsou přítomné i binární příznaky říkající například zda je daný den víkend (Obrázek 2.4).



Obrázek 2.2: Ukázka sinové a kosinové transformace na časových příznacích



Obrázek 2.3: Ukázka sinové a kosinové transformace na časových příznacích



Obrázek 2.4: Ukázka binárních a dalších časových příznaků

2.3.11 Sjednocení a časová rozlišení dat

Vzhledem k tomu, že různé zdroje pracují s odlišným vzorkováním (např. 60 minut pro meteorologická data či vybrané energetické řady, 15 minut pro většinu energetických řad), je nezbytným krokem sjednocení časové osy. V rámci této práce bylo zvoleno cílové rozlišení 15 minut, které odpovídá obchodní jednotce v ČR. Pro doplnění chybějících hodnot u hodinových dat byla využita lineární interpolace (metoda upsamplingu), která zajišťuje hladký přechod mezi hodinovými stavy a zabraňuje vzniku schodovitých diskontinuit v datech. Tabulka 2.1 přehledně shrnuje dostupnost a charakteristiku všech využitých veličin.

Tabulka 2.1: Souhrnný přehled dostupnosti a technických parametrů datových zdrojů.

Kategorie	Veličina	Typ	Čas dostupnosti	Rozlišení
Energetika	Fyzikální toky	Skutečné	Průběžně	15 min
	Celková výroba	Skutečné	Průběžně	15 min
	Celková výroba	Predikce	D-1 (18:00)	15 min
	Obchodní výměny	Plánované	D-1 (13:00)	15 min
	Předpověď zatížení	Predikce	D-1 (10:00)	15 min
	Předpověď OZE (Zítra)	Predikce	D-1 (18:00)	15 min
	Předpověď OZE (Intra)	Predikce	D+0 (8:00)	15 min
	Tržní ceny další den	Predikce	D-1 (13:00)	15 min
Počasí	Reanalýza ERA5	Skutečné	D+5	60 min
	Historická předpověď	Predikce	Průběžně	60 min
	Předpověď počasí	Predikce	až D-14	60 min
Ostatní	Kalendářní data	Determin.	-	-

2.4 Metody úpravy dat

Hrubá data získaná z API portálů nelze přímo aplikovat do modelů strojového učení. Je nutné provést nejprve předzpracování dat:

2.4.1 Agregace dat

Knihovna `entsoe-py` [Ene26] dokáže poskytnout například data pro výrobu rozdělená podle jednotlivých výrobních zdrojů. Tyto hodnoty jsou pro účely predikce celkového tranzitu agregovány do souhrnných řad, čímž dochází ke snížení dimenzionality vstupních dat.

2.4.2 Výpočet salda

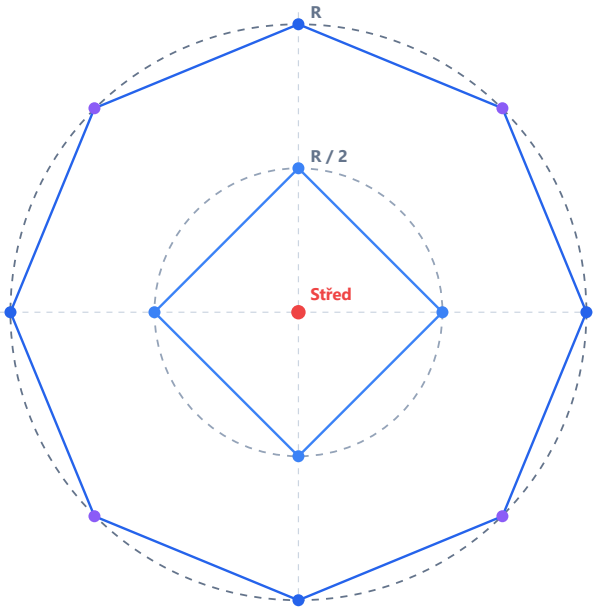
Aby model dokázal lépe vnímat celkovou pozici daného státu či profilu, jsou pro vybrané časové řady (řady týkající se hranic) počítány čisté hodnoty. Ty se vypočítávají přímo odečtením importní hodnoty od exportní. Výsledná hodnota a její znaménko jednoznačně indikují, zda je daná země/hranice v daném časovém úseku čistým vývozcem či dovozcem energie.

2.4.3 Rozdíl cen

Fyzikální tok energie souvisí s ekonomickým modelem evropského trhu. Z hlediska predikce má vyšší informační hodnotu než absolutní cena elektřiny v jedné zemi právě cenový rozdíl mezi sousedícími obchodními zónami. Nízká cena v jedné zemi a vysoká v zemi druhé indikuje silný komerční zájem transportovat energii mezi zeměmi, což se následně projeví na fyzických přeshraničních tocích. Do datasetu jsou proto pomocí funkce dopočítávány diference denních cen pro uživatelem specifikované páry zemí.

2.4.4 Prostorové průměrování hodnot počasí

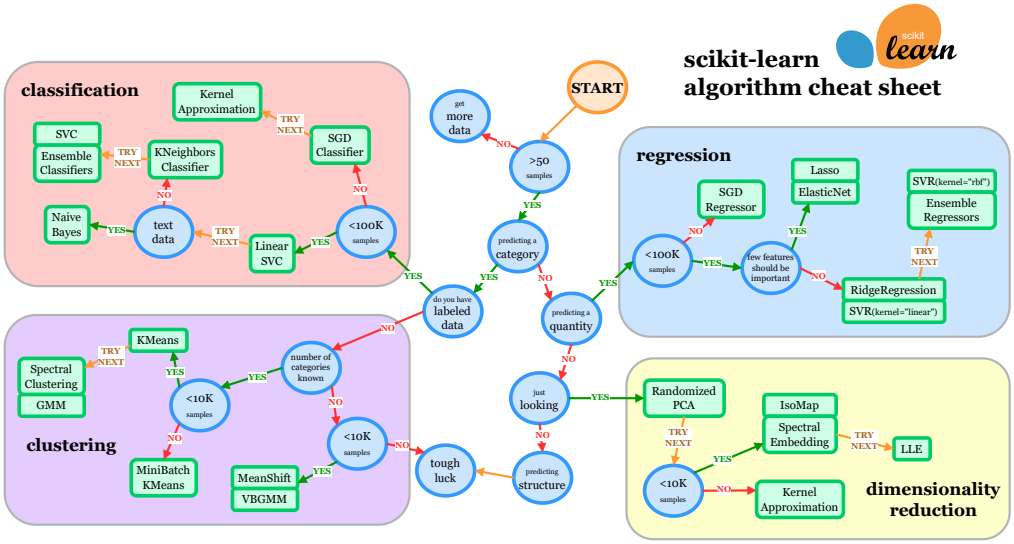
Uvažování teploty či rychlosti větru pouze z jednoho konkrétního bodu může být pro modely zavádějící. Používá se proto prostorové průměrování dat z několika měřicích bodů rozprostřených do specifického polygonu pro zajištění konzistence. Všech 13 bodů má v tomto případě nastavenou stejnou váhu. Metoda prostorového průměrování umožňuje posbírat data z bodů definujících určitou geografickou oblast, čímž vzniká vyhlazenější a mnohem přesnější agregovaná metrika počasí pro celou velkou oblast či stát.



Obrázek 2.5: Vzorec použitý pro průměrování

2.5 Základní vybrané metody modelování

Pro predikci tranzitu byly zvoleny prediktivní algoritmy určené pro regresní úlohy. Zvolené algoritmy byly vybrány napříč různými kategoriemi s cílem otestovat odlišné přístupy k regresi. Specificky byl také použit graf z webových stránek `scikit-learn` pro výběr modelů [Ped+26]. Ten doporučuje pro menší až středně velké datasety s lineární či mírně nelineární závislostí začít od základních a dobře interpretovatelných modelů. Metody hlubokého učení a neuronových sítí nebyly v této práci použity, většina práce spočívala v úpravě a přípravě datasetu. Dále popsané regresní metody byly vyhodnoceny jako dostatečné.



Obrázek 2.6: Rozhodovací proces při výběru modelu podle scikit-learn. Zdroj: [Ped+26]

2.5.1 `LinearRegression`, `Lasso`, `Ridge`

Lineární regrese tvoří absolutní základ a slouží jako jednoduchý model, se kterým se porovnávají složitější modely. Tento model předpokládá lineární závislost mezi cílovou proměnnou a vstupy. Je výpočetně extrémně nenáročný. Problémem klasické lineární regrese v datasetech s velkým množstvím příznaků je ale hrozba přeučení. Z toho důvodu je vhodné využít varianty pracující s penalizací složitosti modelu – např. algoritmy `Lasso` (L_1 regularizace) nebo `Ridge` (L_2 regularizace). Zatímco `Ridge` váhy nedůležitých parametrů efektivně potlačuje, `Lasso` je dokáže zcela vynulovat, čímž samo provádí selekci příznaků.

2.5.2 `LinearSVR`

Metoda podpůrných vektorů pro regresi se nesnaží pouze minimalizovat klasickou chybu, ale snaží se minimalizovat i "hranice" funkce, čímž se zaměřuje na maximalizaci obecnosti. `LinearSVR` zavádí epsilon pásmo tolerované chyby; pokud predikce padne do tohoto pásma, není penalizována. Tento algoritmus je robustní vůči odlehlým hodnotám, které se v energetických datech často objevují v dobách nečekaných výpadků v síti nebo extrémních výkyvů počasí. Oproti klasické SVR s nelineárním jádrem je `LinearSVR` výpočetně efektivnější pro datasety s velkým množstvím příznaků.

2.5.3 ElasticNet

Tento model představuje kompromis mezi regresními algoritmy `Lasso` a `Ridge`. `ElasticNet` kombinuje obě formy penalizace (L_1 a L_2 regularizaci), což mu umožňuje těžit z výhod obou přístupů. V kontextu predikce energetických toků je tento model užitečný zejména v případech, kdy existuje silná korelace mezi jednotlivými příznaky, se kterou si samotné `Lasso` nemusí poradit optimálně. `ElasticNet` dokáže selektovat relevantní skupiny proměnných při zachování stability vah, což je klíčové při integraci heterogenních dat o počasí a výrobě z různých zemí.

2.5.4 SGDRegressor

Algoritmus `SGDRegressor` implementuje lineární regresi využívající stochastický gradientní sestup. Tento přístup je vysoce efektivní pro trénování modelů na rozsáhlých datasetech, protože optimalizace probíhá postupně po jednotlivých vzorcích nebo malých dávkách. V predikčním systému pro energetiku je `SGDRegressor` vhodný díky své schopnosti pracovat s velkým množstvím příznaků při relativně nízkých paměťových nárocích. Model je rovněž citlivý na škálování vstupních dat, což v implementaci řeší proces normalizace v rámci trénovací pipeline.

2.5.5 HistGradientBoostingRegressor

Na rozdíl od předchozích lineárních modelů je zastoupena i moderní metoda založená na souborovém učení. `HistGradientBoostingRegressor` je založen na rozhodovacích stromech. Algoritmus je inspirován architekturou LightGBM a využívá shlukování (rozdělení spojitých hodnot do histogramů) pro výrazné zrychlení procesu učení na velkých objemech dat. Pro úlohu predikce tranzitu je tento model zásadní svou schopností automaticky zachycovat nelineární vztahy a komplexní interakce mezi příznaky (např. vliv nárazového větru v kombinaci s aktuální cenou elektřiny), které čistě lineární modely nemusí plně postihnout.

2.6 Metody vyhodnocení

Vyhodnocení prediktivního modelu na časových řadách vyžaduje specifický přístup, protože nelze provést standardní náhodné rozdělení na trénovací a testovací data. Došlo by k nežádoucímu úniku informací z budoucnosti. Data tedy byla rozdělena podle roků měření, roky 2023 a 2024 jsou používané pro trénování modelu, rok 2025 následně slouží pro jeho validaci pomocí metrik. Dále bylo použito ještě období od 1. 1. 2026 do 31. 3. 2026 pro testovací účely. Testování v reálném provozu probíhalo od 20. 4. 2026 do 14. 5. 2026. Rozdělení po celých letech dává modelu dobré předpoklady pro zachycení i sezónních jevů.

2.6.1 Metriky vyhodnocování

Pro kvantitativní posouzení kvality predikčních modelů byla zvolena sada regresních metrik. V níže uvedených vztazích jsou použity následující proměnné:

- n – celkový počet vzorků v testované sadě,
- y_i – skutečná naměřená hodnota (ground truth) v čase i,
- $\hat{y}_i$ – hodnota predikovaná modelem pro daný čas i,
- $\bar{y}$ – aritmetický průměr skutečných hodnot v celé sadě,
- $\bar{\hat{y}}$ – aritmetický průměr všech predikovaných hodnot v sadě.

1. **MAE (Mean Absolute Error):** Vyjadřuje průměrnou absolutní odchylku predikce od skutečné hodnoty. Výhodou je zachování jednotek (v tomto případě MW) a snadná interpretovatelnost pro dispečery sítě.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.1}$$

2. **MAPE (Mean Absolute Percentage Error):** Vyjadřuje chybu v procentech, což umožňuje porovnávat úspěšnost predikcí mezi hranicemi s naprosto odlišnými absolutními objemy protékající energie. Tato metrika ovšem není vhodná pro případy, kdy se hodnota často pohybuje okolo nulové hodnoty.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{2.2}$$

3. **RMSE (Root Mean Squared Error):** Odmocnina průměrné kvadratické chyby. Tato metrika je vysoce senzitivní na extrémně velké chyby. Nízké RMSE indikuje, že se model vyhýbá velkým odchylkám, což je pro bezpečnost kritické energetické infrastruktury klíčové.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.3}$$

4. **R^2 (Koeficient determinace):** Statistická míra určující podíl rozptylu závislé proměnné, který je modelem predikován z nezávislých proměnných. Nabývá hodnot maximálně 1 (dokonalá shoda) a ukazuje relativní úspěšnost modelu vůči predikci pouhým průměrem.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.4}$$

5. **Korelační koeficient (Pearsonův):** Udává míru lineární závislosti mezi predikovanými a skutečnými hodnotami. Hodnota blízká 1 značí, že trend predikce velmi přesně odpovídá trendu skutečných dat, což je důležité pro predikci směru toku energie. Využívá se také v Kapitole 4.4, kde se počítá korelace mezi predikovanou řadou a jednotlivými příznaky.

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{2.5}$$

Návrh a implementace

3

Pro jednodušší stahování dat byly vytvořeny třídy přistupující k příslušným API. Hlavními třídami jsou `EnergyLoader`, `WeatherLoader` a `FeatureLoader`. Každá třída načítá vlastní datovou strukturu, která je podtřídou `TimeSeries`. V této třídě jsou data uložena především ve formátu Pandas `DataFrame`. Třídy umožňují provádět operace nad dostupnými daty. Dále je architektura rozdělená do jednotlivých kroků, kde data procházejí třídami jako `DatasetPreparer`, `ModelTrainer` a `ModelPredictor`. Celá architektura je vyobrazena v Diagramu 3.1.

Veškeré zdrojové kódy vytvořené v rámci této práce jsou veřejně dostupné v repozitáři na platformě GitHub[1].

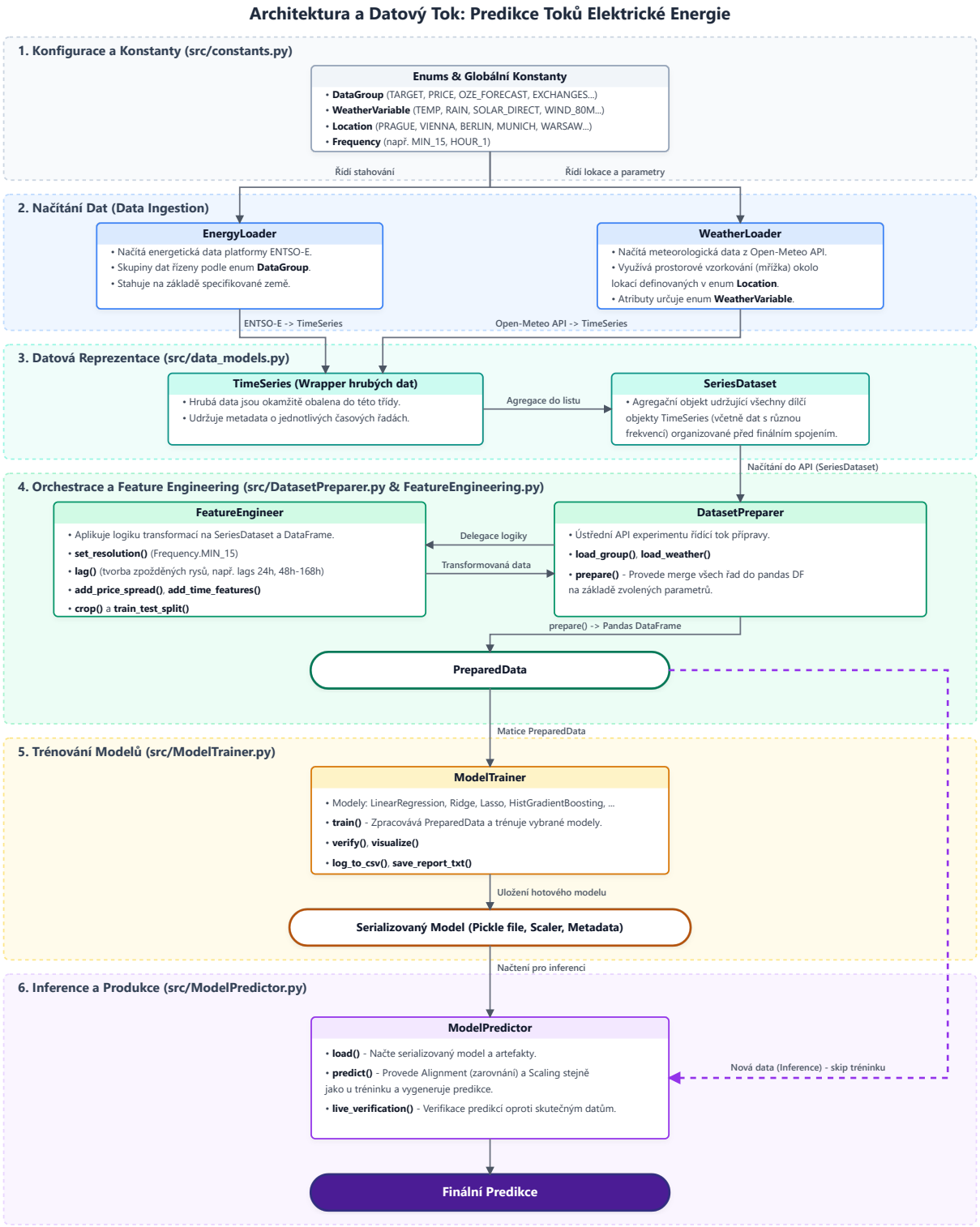
3.1 Implementace softwarového řešení

Tato kapitola popisuje technickou realizaci softwarových nástrojů vyvinutých pro sběr, harmonizaci a správu dat. Vzhledem k heterogenní povaze vstupních dat (energetické časové řady vs. meteorologická prostorová data) a nutnosti zpracovat velké objemy informací z více zdrojů, byl kladen důraz na modulární a objektově orientovanou architekturu v jazyce Python. Řešení je rozděleno do tří logických vrstev:

1. **Datová vrstva:** Definuje jednotné struktury pro ukládání dat.
2. **Integrační vrstva:** Zajišťuje komunikaci s externími API.
3. **Aplikační vrstva:** Skripty pro hromadné stahování, trénování a validaci.

[1] `https://github.com/Lendoran/BP`

3.2 Diagram projektu



Obrázek 3.1: Diagram projektu predikce tranzitu

3.3 Načítání z datových zdrojů

3.3.1 Dynamické načítání

Získávání dat je primárně řízeno třídou `DatasetPreparer`, která funguje jako hlavní orchestrátor, a specializovanými loadery (`EnergyDataLoader` a `WeatherLoader`). Načítání probíhá dynamicky – systém nejprve zkontroluje, zda jsou požadovaná data již k dispozici lokálně. V experimentálním či trénovacím režimu, kde jsou potřeba data mimo lokální úložiště, se chybějící data automaticky stáhnou a podle nastavení i uloží. Pro nasazení v reálném čase je implementován "živý mód", kdy jsou pomocí časových parametrů data stažena přímo do paměti a proces obchází lokální úložiště, čímž je zajištěna práce s co nejaktuálnějšími informacemi. K zamezení zbytečných volání API při výpadku zdrojů slouží interní třída `DownloadTracker`, která neúspěšné pokusy loguje a problematické řady zařazuje na seznam blokovaných zdrojů.

3.3.2 Ukládání trénovacích dat

Pro zrychlení opakovaného běhu a snížení zátěže externích serverů je implementován systém ukládání. Stažená a zpracovaná data z energetického sektoru jsou serializována pomocí knihovny `pickle` a ukládána do adresářové struktury definované konstantou. Tento formát byl zvolen pro svou nativní podporu komplexních Python objektů a rychlost zápisu během experimentální fáze. U meteorologických dat, kde dochází k masivnímu dotazování, je na úrovni síťové komunikace implementována mezipaměť (cache) přímo pomocí knihovny `requests_cache`, která po definovanou dobu expirace vrací lokální odpovědi na identické API dotazy.

3.3.3 Využití ENUM konstant

Pro zajištění typové bezpečnosti, prevence překlepů a celkové standardizace kódu jsou klíčové parametry systému definovány pomocí výčtových typů (Enum) ve sdíleném souboru `constants.py`. Zahrnují proměnné datových sad (`DataGroup`, např. `TARGET`, `PRICE`, `GENERATION`), mapování meteorologických proměnných (`Weather Variable`, např. `TEMP`, `WIND_80M`) a geografické lokality (`Location`). Tento přístup umožňuje snadné iterování přes skupiny dat při hromadném načítání a usnadňuje logiku v hlavním kontroleru. Konstanty jsou dostupné na odkazu [2]

[2] `https://github.com/Lendoran/BP/blob/main/src/constants.py`

3.3.4 Komunikace s ENTSO-E API

Komunikace s platformou ENTSO-E je zapouzdřena ve třídě `EnergyDataLoader`, která využívá wrapper `EntsoePandasClient`. Vzhledem k přísným limitům platformy na objem stahovaných dat v jednom dotazu a častým timeoutům u rozsáhlých časových řad byla implementována metoda `_fetch_chunked()`. Ta obdržený časový interval (např. 3 roky) automaticky rozdělí na roční bloky a stahuje je postupně. Metoda obsahuje detailní logování stavu v konzoli a v případě chyby udrží běh programu s tím, že spojí úspěšně stažené bloky a ošetří duplicitní indexy.

3.3.5 Komunikace s Open-Meteo API

Meteorologická data jsou získávána přes Open-Meteo API a třídu `WeatherLoader`. Z důvodu dodržení limitů bezplatného přístupu je asynchronní komunikace ošetřena validací a parametrem `backoff_factor` v rámci knihovny `retry_requests`. Skript navíc disponuje logikou, která na základě stáří požadovaných dat dynamicky přepíná mezi URL adresami (Live Forecast API pro nedávnou a budoucí dobu, Historical Forecast API pro archiv).

3.4 Příprava datasetu

3.4.1 Pipeline proces

Surová data získaná z loaderů se shromažďují v instanci třídy `SeriesDataset`, která uchovává objekty typu `TimeSeries`. Třída `DatasetPreparer` následně slouží jako pipeline, přes kterou jsou postupně volány metody pro čištění, harmonizaci a rozšiřování dat. Metody vrací instanci sama sebe, což umožňuje řetězení příkazů. Každý objekt `TimeSeries` si sebou nese nejen samotná data, ale i svá metadata – jednotky, frekvenci vzorkování a zpoždění dostupnosti.

3.4.2 Načítací skupiny

Načítání do datové struktury je rozděleno do logických skupin (`load_group`), přičemž každá skupina (např. `TARGET`, `GENERATION`, `LOAD`) z konstant prochází specifickým rozbalením (`_expand_and_add_dataset()`). Pro zajištění kompatibility s modely strojového učení se vícerozměrné datasety z API (například výroba rozdělená podle typu elektrárny) transformují na jednodimenzionální časové řady, z nichž každá reprezentuje jeden samostatný příznak a je zařazena pod vlastním názvem do slovníku sady.

3.4.3 Úpravy datasetu

Sjednocení parametrů datové sady zajišťuje pomocná třída `FeatureEngineer`. Mezi hlavní úpravy patří nastavení jednotného rozlišení funkcí `set_resolution()` (např. vše na hodinový interval pomocí průměrování), ořezání okrajových dat (metoda `crop()`), či imputace (doplňování) chybějících hodnot metodou `fill_gaps()`. Zásadní součástí je vytváření zpožděných hodnot (`add_lag()`), které generují nové časové řady posunuté o zadaný počet hodin (např. 24, 168 pro denní a týdenní sezónnost). Všechny aktuálně dostupné datové řady a jejich pokrytí lze vypsat pomocí zabudovaných monitorovacích funkcí `info()` a `print_available_series()`.

3.4.4 Další metody úprav

Prostorový dopad počasí je řešen metodou prostorového průměrování. Kolem zadané geografické souřadnice se vytvoří 13 deterministických bodů ve tvaru čtverce a osmiúhelníku (Obrázek 2.5). Výsledky z těchto bodů jsou následně v loaderu zprůměrovány, což vytváří mnohem stabilnější prediktor než měření z jediného uzlu. Z teplotních křivek jsou dopočítávány odvozené míry – hodiny vytápění (HDH) a hodiny chlazení (CDH). Tyto hodnoty se počítají jako počet hodin pod a nad referenční teplotou, kdy HDH představuje sumu stupňohodin pod 18 °C a CDH sumu nad 22 °C. Pro implementaci jakýchkoliv arbitrárních vlastních výpočtů (pomocí Lambda funkcí či externích metod) slouží robustní metoda `add_custom_feature`. Ta využívá interní objekt `MetadataTracker`, který sleduje, z jakých sloupců nová funkce (např. lambda operace) čerpá data, a na základě těchto informací automaticky přiřadí nejhorší možné zpoždění dostupnosti finálnímu příznaku, čímž preventivně eliminuje data leakage.

3.5 Trénování modelů a vyhodnocení

Proces učení řídí třída `ModelTrainer`. Třída natrénuje nejdříve model pro nulový horizont, poté využívá časový posun celého bloku příznaků vůči cílové proměnné. Znamená to, že pro predikci na X hodin dopředu jsou používány proměnné z predikovaného času, které jsou již dostupné v čase predikce. Samotná implementace algoritmu je plně abstrahována, metoda `train()` přijímá odkaz na libovolnou třídu modelu ze `scikit-learn` knihovny a asociativní pole argumentů (**kwargs). Lze tak v rychlém sledu iterovat a testovat modely typu `LinearRegression`, `Lasso`, `LinearSVR` či jiné bez nutnosti zasahovat do struktury trenéru.

3.5.1 Vyhodnocovací metody

Po natrénování modelu jsou ověřovány jeho schopnosti na vyčleněném testovacím vzorku metodou `verify()`. Ta vrací asociativní pole klíčových regresních metrik – kromě základní absolutní chyby (MAE) počítá odmocninu střední kvadratické chyby (RMSE), chybu v procentech (MAPE) implementovanou s maskou ošetřující dělení nulou, koeficient determinace (R^2) a hodnotu korelace (Corr) mezi predikcemi a reálným tokem.

3.5.2 Zobrazování výsledků

Veškeré výsledky se agregují v reálném čase. Funkce `log_to_csv()` zajišťuje průběžné uložení sady výsledků do souhrnného souboru `experiments_summary.csv`, což umožňuje pozdější detailní srovnávání. Systém generuje také podrobné textové rozpisy obsahující seznam a váhy uplatněných příznaků pro daný test. Průběh programu ukládá logy pomocí funkcí `_log()` a `_log_raw()` do dedikovaného textového souboru. Vizuální validace zajišťuje metoda `visualize()`, jež kreslí překryv predikce vůči skutečnosti v časovém okně společně s rámečkem chybových metrik a generuje obrázky v neblokujícím grafickém okně.

3.6 Přechod na živé predikce

Systém je navržen s ohledem na možnost fyzického nasazení. Třída `ModelTrainer` disponuje funkcí `export_models()`, která po dokončení trénování serializuje celý slovník výsledků (obsahující natrénovanou instanci modelu, pole nutných sloupců, stav scaleru pro normalizaci dat a další) do artefaktového souboru typu .pickle. V produkčním prostředí se následně instancuje lehká třída `ModelPredictor`, která pomocí metody `load()` tento artefakt načte. Tím se odděluje predikční fáze od trénovací a nevzniká přímá závislost na struktuře trénování modelu nebo původním vývojovém prostředí.

3.6.1 Průběžné stahování hodnot z API

Pro otestování úspěšnosti predikcí v reálném nasazení byl vytvořen skript pro stáhnutí všech dostupných dat z API. Tento skript byl spouštěn každou hodinu a ukládal aktuálně dostupné hodnoty jako uloženou třídu `DatasetPreparer` a také jako csv tabulku pro jednodušší vizualizování.

3.6.2 Čas predikce

Třída `ModelPredictor` obsahuje metodu `predict()`, která generuje na základě parametrů předpověď pro daný časový úsek (například horizonty H+0 až H+14). Při nasazení vzniká na základě různých aktualizací dat rozdíl v použitých hodnotách pro predikci. Predikce provedená např. v 10:00 může poskytovat částečně jinou množinu budoucích bodů než predikce ve 22:00. Toto je případ hlavně dat o počasí, která jsou aktualizována i v průběhu dne, proto byly vybrány dané časy predikcí. V 10:00 jsou dostupné aktualizované hodnoty pro daný den, predikuje se pouze vnitrodenně. V 22:00 jsou dostupná všechna energetická data pro následující den, je tedy možné predikovat i hodnoty z dalšího dne.

Výsledky a vyhodnocení

4

Pro porovnání výkonnosti modelů a vlivu jednotlivých typů dat byla zvolena metoda vyčerpávajícího testování všech kombinací definovaných příznakových sad. Vstupní příznaky byly rozděleny do osmi disjunktních tříd, které reprezentují různé aspekty predikčního problému:

1. **Třída B (Baseline)**: obsahuje základní srovnávací prvek, kterým je hodnota cílové veličiny zpožděná o 24 hodin (hodnota z předchozího dne ve stejný čas).

2. **Třída A (Autoregresivní)**: zahrnuje hodnoty cílové veličiny se zpožděním 25 až 168 hodin. To modelu umožňuje zachytit týdenní sezónnost a dlouhodobější závislosti v datech. Tato třída ale násobně rozšiřuje celou datovou sadu a zpomaluje trénování modelů. Po korelační analýze a testování bylo vyhodnoceno jako vhodnější použít pouze nejpřínosnější hodnoty, a to ty se zpožděním o 48, 72, 96, 120, 144 a 168 hodin.

3. **Třída W (Počasí)**: obsahuje meteorologické údaje, jako jsou teplota, vlhkost vzduchu či intenzita solárního záření.

4. **Třída C (Lokální energetika)**: zahrnuje specifické hodnoty pro Českou republiku, zejména predikce výroby z obnovitelných zdrojů, plánované přeshraniční výměny, očekávanou spotřebu a celkovou výrobu.

5. **Třída N (Sousední země)**: obsahuje obdobné energetické veličiny jako třída C, avšak pro zemi, se kterou je predikován tok na společné hranici.

6. **Třída E (Ostatní sousedé)**: doplňuje energetické hodnoty pro zbývající sousední státy v rámci propojené soustavy.

7. **Třída T (Časové příznaky)**: reprezentuje časové údaje transformované pomocí funkcí sinus a kosinus (cyklické kódování), což modelu usnadňuje interpretaci periodických jevů.

8. **Třída S (Speciální hodnoty)**: obsahuje odvozené veličiny, například cenové diferenciály mezi jednotlivými tržními oblastmi.

Kombinací těchto osmi tříd vzniká celkem $2^8 = 256$ možných datových sad. Po vyloučení prázdné sady bylo pro každý model testováno 255 validních variant vstupních dat. Samotný výběr modelů pro testování reflektoval doporučení knihovny `scikit-learn` pro daný typ úlohy, jak bylo uvedeno dříve u Obrázku 2.6. Celkem bylo vyhodnoceno sedm modelů zastupujících kategorie lineárních modelů, SVM a ensemble modelů.

4.1 Výsledky pro jednotlivé hraniční profily

V této podkapitole je provedeno detailní srovnání výkonnosti modelů na jednotlivých přeshraničních profilech CZ→(AT, DE, PL, SK). Pro každý profil je porovnána skutečně naměřená hodnota se třemi stupni predikce:

- **Základní model (Baseline):** Využívá výhradně příznakovou třídu B. Jako algoritmus je použita standardní lineární regrese.
- **Autoregresní model:** K základní třídě přidává rozšířené autoregresní příznaky (třídy BA), modelem je opět lineární regrese.
- **Nejlepší model:** Reprezentuje nejúspěšnější nalezenou kombinaci algoritmu a datové sady vzešlou z vyčerpávajícího testování pro daný konkrétní profil.

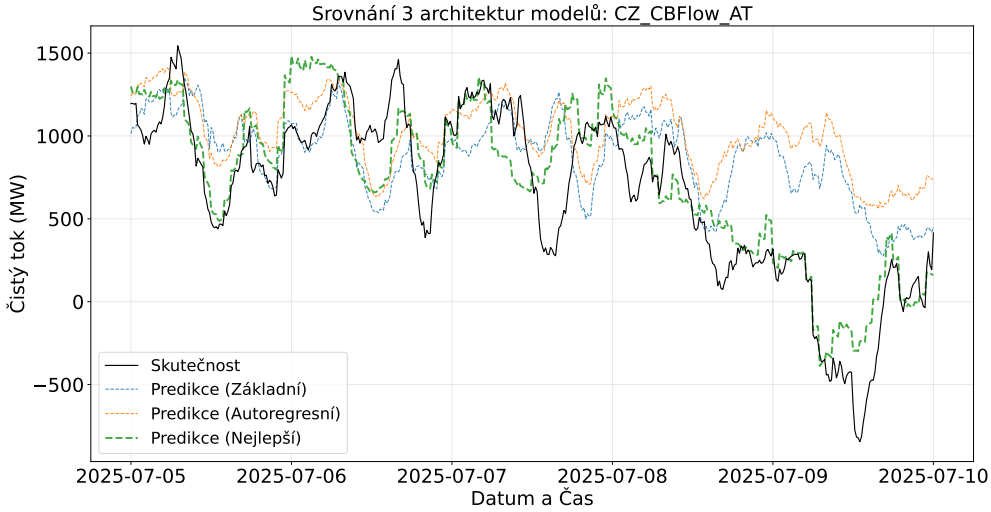
Vizuální reprezentace vybraného časového úseku (od 5. do 10. července 2025) ilustruje schopnost modelů reagovat na dynamické změny v síti.

4.1.1 Profil CZ → AT

Na hranici s Rakouskem dosáhl nejlepších výsledků model typu `ElasticNet` natrénovaný na datech z příznakových tříd BAWCNE. Skutečné fyzikální toky na tomto profilu v analyzovaném období oscilovaly převážně v rozmezí přibližně −500 až 1500 MW (Obrázek 4.1).

Základní model s jedinou příznakovou skupinou B dosahoval průměrné absolutní chyby (MAE) 310,20 MW s koeficientem determinace R^2 pouze 0,5121. Přidání autoregresních příznaků A chybu mírně snížilo na 285,45 MW. Až nasazení nejlepšího modelu `ElasticNet` s využitím informací o počasí, lokální energetiky a sousedních zemí dokázalo MAE stlačit na 204,77 MW a signifikantně zvýšit schopnost modelu vysvětlit rozptyl dat ($R^2 = 0,7854$).

Z grafického průběhu (Obrázek 4.1) je jasně patrné, že základní model je v podstatě pouze fázově posunutou kopií skutečnosti o 24 hodin, což způsobuje masivní chyby při náhlých změnách trendu. Nejlepší model naopak dokáže mnohem přesněji sledovat skutečnou trajektorii i při náhlých propadech.

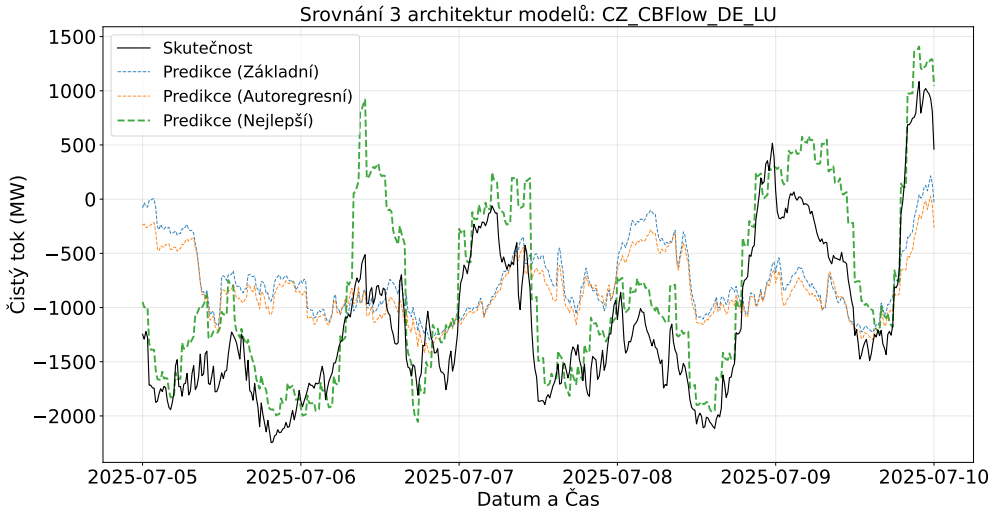


Obrázek 4.1: Srovnání predikcí s reálnými toky na profilu CZ → AT

4.1.2 Profil CZ → DE_LU

Profil s Německem představuje z hlediska predikce nejnáročnější relaci, vyznačující se extrémní volatilitou s propady až k −2000 MW a špičkami okolo 1000 MW (Obrázek 4.2). Nejlepších výsledků zde dosáhl model `SGDRegressor` s využitím sady BWCNTE. Zajímavostí je, že tento model vynechal autoregresní třídu, ale využil časové příznaky, údaje o počasí a energetická data okolních zemí.

Výchozí MAE základního modelu činilo 735,19 MW s velmi nízkým koeficientem $R^2 = 0,2972$, což naznačuje, že extrapolace toku pouze z předchozího dne je pro německou hranici zcela nedostačující. Výkonnost se zásadně změnila u nejlepšího modelu, který snížil MAE na 312,29 MW a zvýšil R^2 na 0,8630. Z grafu na Obrázku 4.2 je vidět, že základní modely zcela selhávají při predikci prudkých poklesů, zatímco nejlepší model s těmito extrémními fluktuacemi dokáže poměrně dobře držet krok.

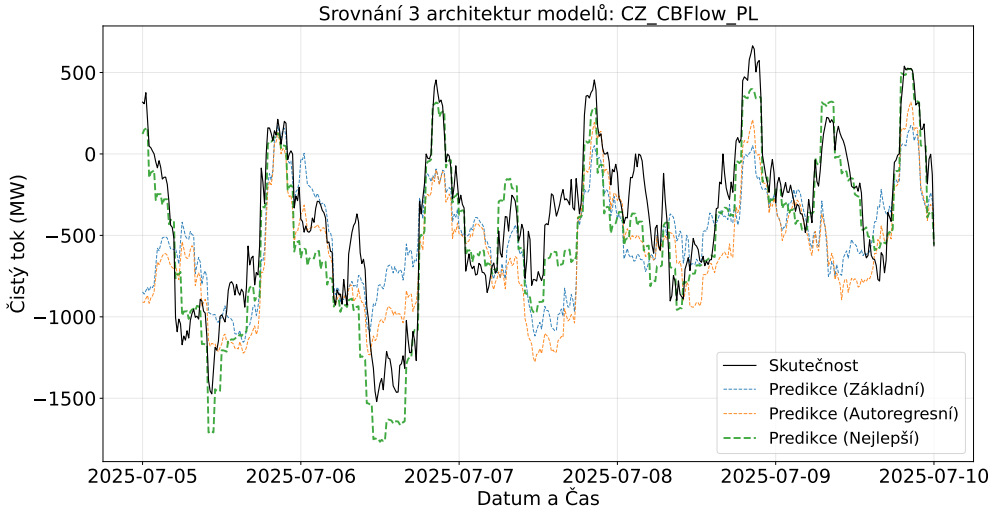


Obrázek 4.2: Srovnání predikcí s reálnými toky na profilu CZ → DE_LU

4.1.3 Profil CZ → PL

Polský profil se vyznačuje toky pohybujícími se převážně v záporných hodnotách v rozmezí zhruba −1500 až 500 MW (Obrázek 4.3). Z vyčerpávajícího prohledávání vzešel jako vítěz model `LinearSVR` využívající datovou sadu BAWNTE.

Kvantitativní metriky ukazují posun od MAE 416,56 MW u základního modelu přes 392,89 MW u autoregresního přístupu až k hodnotě MAE 188,95 MW u modelu nejlepšího. Tím došlo ke zlepšení koeficientu determinace na 0,8617 a celková korelace se skutečností dosáhla hodnoty 0,9329. Vizuální srovnání v grafu na Obrázku 4.3 potvrzuje schopnost nejlepšího modelu precizně kopírovat i lokální maxima a minima v průběhu dne, na které základní perzistence nedokázala tak dobře reagovat.

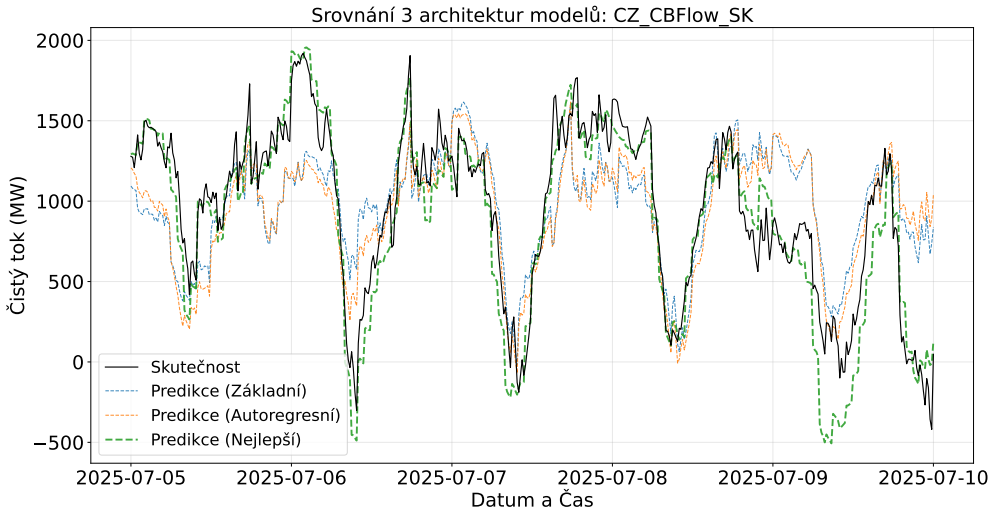


Obrázek 4.3: Srovnání predikcí s reálnými toky na profilu CZ → PL

4.1.4 Profil CZ → SK

Tranzit směrem na Slovensko probíhá typicky v silně kladných hodnotách, často dosahujících hranice i 2000 MW (Obrázek 4.4). Jako nejvhodnější se pro tuto dynamiku profilu ukázal penalizační model `Lasso`, trénovaný na jedné z nejrozsáhlejších příznakových sad: BAWCNTE.

Lineární regrese jakožto základní model dosáhla na tomto profilu MAE 418,78 MW a $R^2 = 0,4679$. Nasazení modelu `Lasso` vedlo ke snížení absolutní chyby na necelou polovinu 207,54 MW a vzestupu vysvětleného rozptylu na hodnotu $R^2 = 0,8674$ s korelací vůči skutečnosti 0,9331. Vykreslený průběh na Obrázku 4.4 demonstruje úspěšné přizpůsobení modelu `Lasso` vysokofrekvenčnímu kolísání toku.

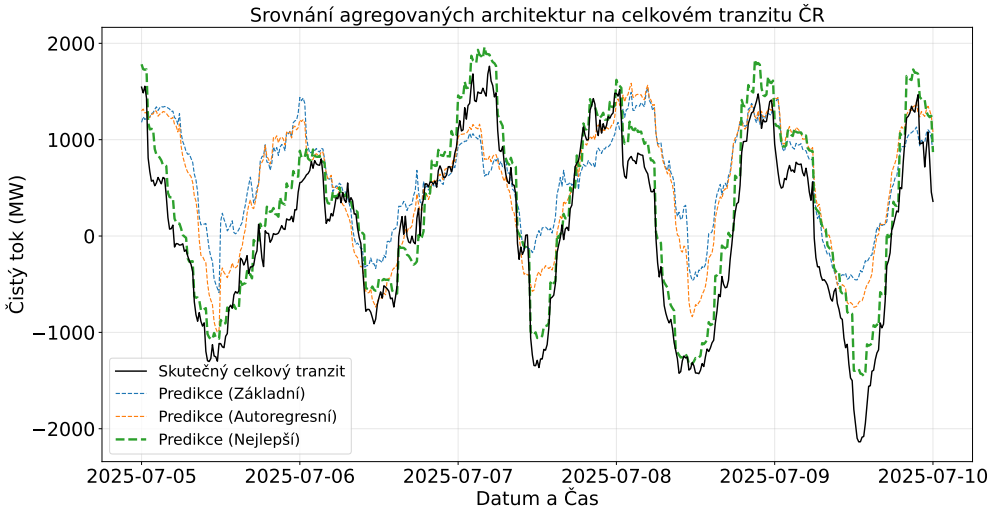


Obrázek 4.4: Srovnání predikcí s reálnými toky na profilu CZ → SK

4.1.5 Celkový tranzit ČR

Klíčovým výstupem práce je predikce celkové tranzitní bilance ČR, definované jako suma predikcí jednotlivých profilů. Tato agregace vykazuje vyšší přesnost díky vzájemné kompenzaci některých lokálních chyb.

Nejlepší kombinovaná architektura dosahuje výsledků s průměrnou chybou MAE 173,52 MW a koeficientem determinace $R^2 = 0,9589$. Jak ilustruje Obrázek 4.5, tento model s vysokou věrností kopíruje skutečný průběh celkové bilance, a to i během prudkých výkyvů. Korelace dosahující hodnoty 0,9794 potvrzuje, že navržený systém velmi přesně aproximuje fyzikální realitu tranzitního uzlu.



Obrázek 4.5: Srovnání predikcí celkového tranzitu ČR

4.1.6 Shrnutí výsledků

Následující Tabulka 4.1 shrnuje metriky pro všechny profily i celkovou bilanci. Je patrné, že využití rozsáhlejších příznakových sad vedlo u všech sledovaných metrik k zásadnímu zlepšení oproti základním modelům.

Tabulka 4.1: Souhrnné porovnání metrik pro jednotlivé hraniční profily a celkový tranzit ČR

Profil	Model (Sady příznaků)	MAE [MW]	RMSE [MW]	MAPE [%]	R^2 [-]	Corr [-]
CZ → AT	`LinearRegression` (B)	310,20	398,35	140,89	0,5121	0,7247
	`LinearRegression` (BA)	285,45	368,21	133,47	0,5831	0,7650
	`ElasticNet` (BAWCNE)	204,77	264,21	98,70	0,7854	0,8877
CZ → DE	`LinearRegression` (B)	735,19	901,80	277,44	0,2972	0,5789
	`LinearRegression` (BA)	701,55	868,55	284,91	0,3481	0,6063
	`SGDRegressor` (BWCNTE)	312,29	398,21	174,21	0,8630	0,9367
CZ → PL	`LinearRegression` (B)	416,56	524,43	297,58	0,3350	0,5874
	`LinearRegression` (BA)	392,89	493,04	293,45	0,4122	0,6439
	`LinearSVR` (BAWNTE)	188,95	239,13	145,91	0,8617	0,9329
CZ → SK	`LinearRegression` (B)	418,78	532,74	166,70	0,4679	0,6981
	`LinearRegression` (BA)	390,32	506,74	166,46	0,5186	0,7238
	`Lasso` (BAWCNTE)	207,54	265,92	73,03	0,8674	0,9331
$\sum CZ$	`LinearRegression` (B)	611,44	762,58	360,14	0,5273	0,7404
	`LinearRegression` (BA)	588,77	739,04	316,63	0,5560	0,7538
	Nejlepší (Suma modelů)	173,52	224,88	111,72	0,9589	0,9794

4.2 Reálné predikce

Tato kapitola se věnuje vyhodnocení úspěšnosti modelů v podmínkách reálného provozu. Testování probíhalo v období od 20. dubna do 14. května 2026. Cílem bylo ověřit robustnost navrženého řešení a jeho schopnost generovat predikce v reálném čase s využitím aktuálně dostupných dat z API rozhraní ENTSO-E a Open-Meteo.

Grafy v této kapitole ilustrují schopnost modelů sledovat skutečný trend fyzikálních toků. Pro každý profil je zobrazen celkový přehled testovaného období a detailní pohled na posledních 7 dní pro lepší čitelnost dynamiky predikce.

4.2.1 Metodika živého testování

Predikce byly generovány pro horizont 14 hodin dopředu, a to dvakrát denně – vždy v 10:00 a ve 22:00. Tento režim odpovídá standardním operačním potřebám, kdy dochází k pravidelné aktualizaci predikcí na základě nově dostupných informací.

Aby byla zajištěna možnost zpětného vyhodnocení, byla data ukládána a načítána způsobem popsaným v Kapitole 3.6.1. Pro účely následné analýzy a verifikace byla v okamžiku každé inferenční operace archivována kompletní stavová data. V případě časového překryvu jednotlivých predikcí byla pro finální vyhodnocení zvolena vždy nejnovější dostupná varianta pro daný okamžik. Tato metodika nejlépe simuluje reálný scénář, ve kterém má dispečer k dispozici nejčerstvější možný odhad vývoje soustavy.

K zajištění maximální věrnosti simulace byly do testování nasazeny nejúspěšnější modely identifikované v Kapitole 4.1.

4.2.2 Kvantitativní výsledky

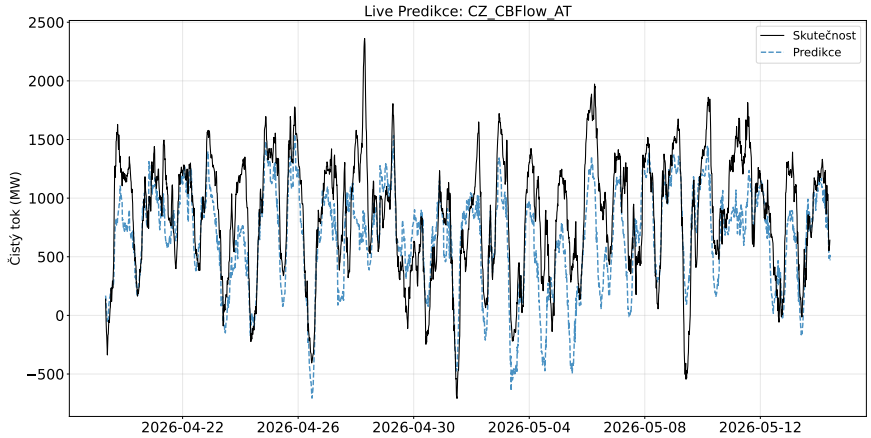
Výsledné metriky dosažené v období reálného provozu jsou shrnuty v Tabulce 4.2. Ve srovnání s výsledky na historických datech (viz Kapitola 4.1.6) je patrný pokles koeficientu determinace R^2 a zhoršení ostatních sledovaných metrik.

Tento jev lze připsat různým faktorům. Jedním z nich je omezený časový rozsah testování. Zatímco modely byly trénovány a validovány na datech pokrývajících celé kalendářní roky, živý test probíhal v relativně krátkém jarním období. Modely tak nemohly plně využít svou schopnost generalizace pro odlišnou sezónní charakteristiku toků, která se projevuje v jiných částech roku.

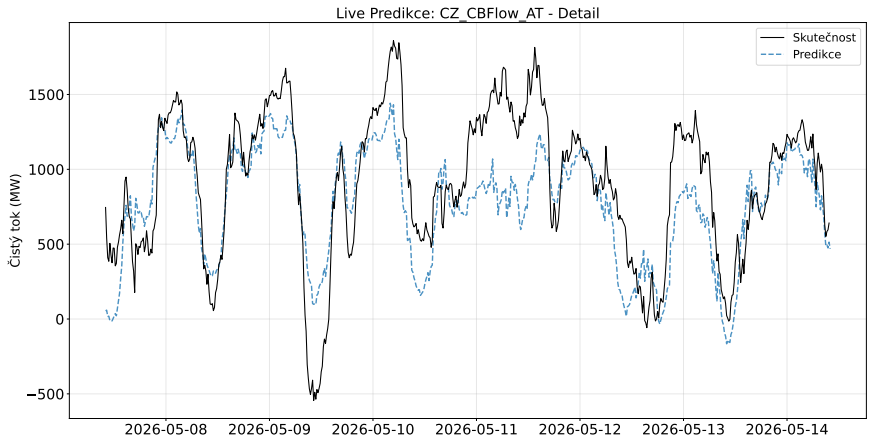
Dalším faktorem je neustále se měnící dynamika energetického mixu. Energetické soustavy evropských států procházejí rychlou transformací s neustálým nárůstem instalovaného výkonu obnovitelných zdrojů energie [Pet26]. Modely trénované na datech z předchozích let tak nemusí plně reflektovat aktuální stav sítě v roce 2026 a zvýšenou volatilitu způsobenou novými zdroji, které v trénovací sadě ještě nebyly zastoupeny.

4.2.3 Profil CZ → AT

U rakouského profilu je patrná stabilní schopnost modelu zachytit základní trend, ačkoliv absolutní chyba MAE (288,38 MW) je v tomto období vyšší než u validace.



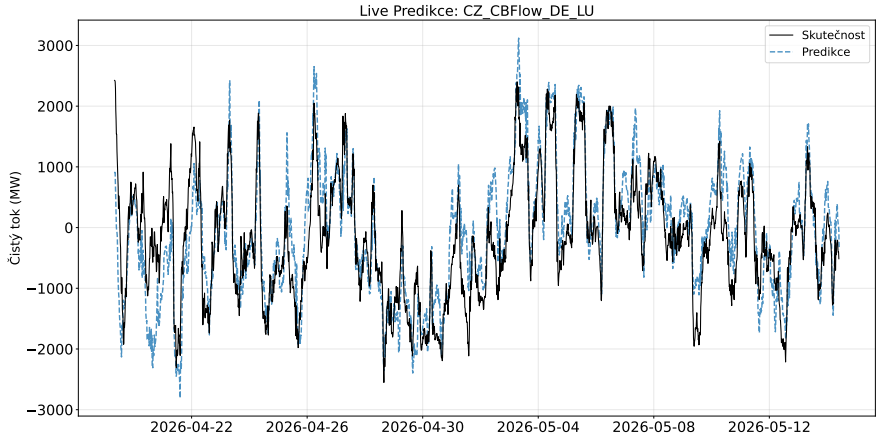
Obrázek 4.6: Celkový průběh reálných predikcí na profilu CZ → AT



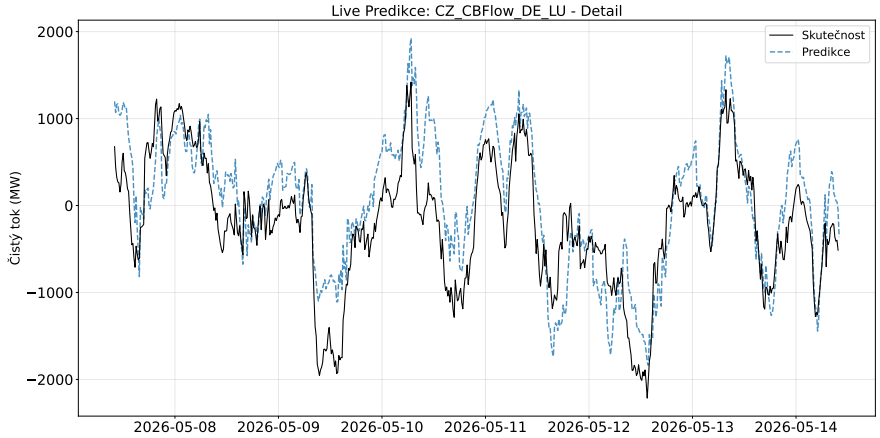
Obrázek 4.7: Detailní pohled na predikci na profilu CZ → AT

4.2.4 Profil CZ → DE_LU

Na německém profilu vykazuje model vysokou korelaci (0,8586) a dokáže dobře sledovat i výrazné propady a špičky toku, což potvrzuje robustnost algoritmu pro tento volatilní trh.



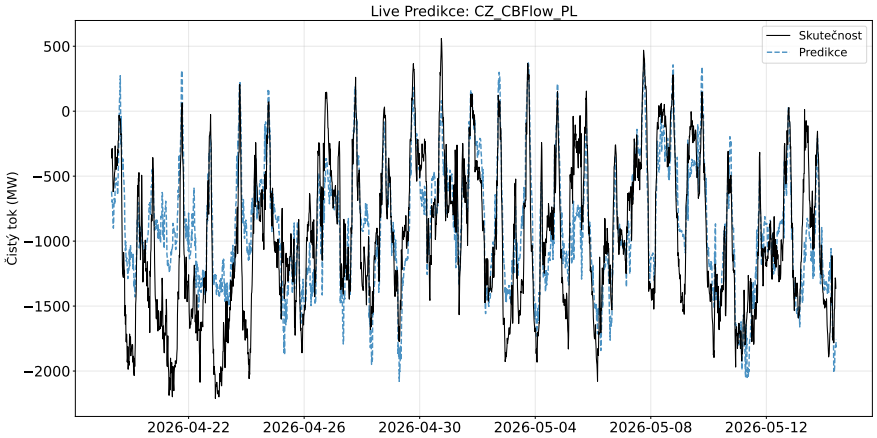
Obrázek 4.8: Celkový průběh reálných predikcí na profilu CZ → DE_LU



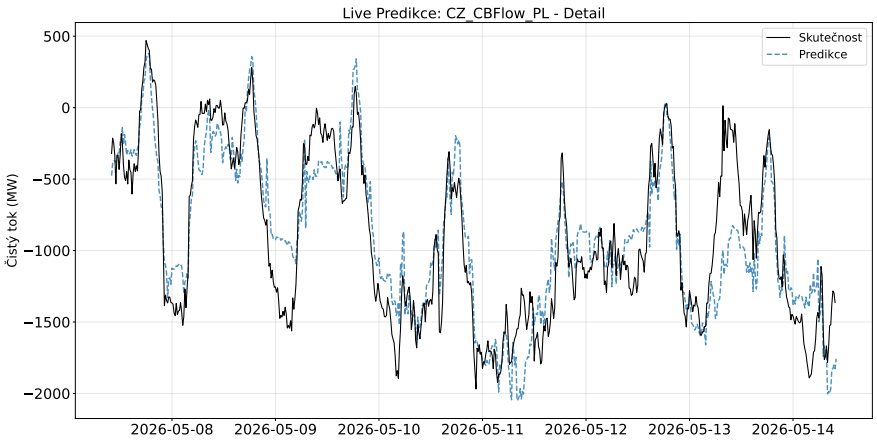
Obrázek 4.9: Detailní pohled na predikci na profilu CZ → DE_LU

4.2.5 Profil CZ → PL

Polský profil dosáhl v reálném provozu absolutní chyby MAE 268,09 MW. I přes jarní volatilitu dokázal model vysvětlit přibližně 64 % rozptylu dat ($R^2 = 0,6391$).



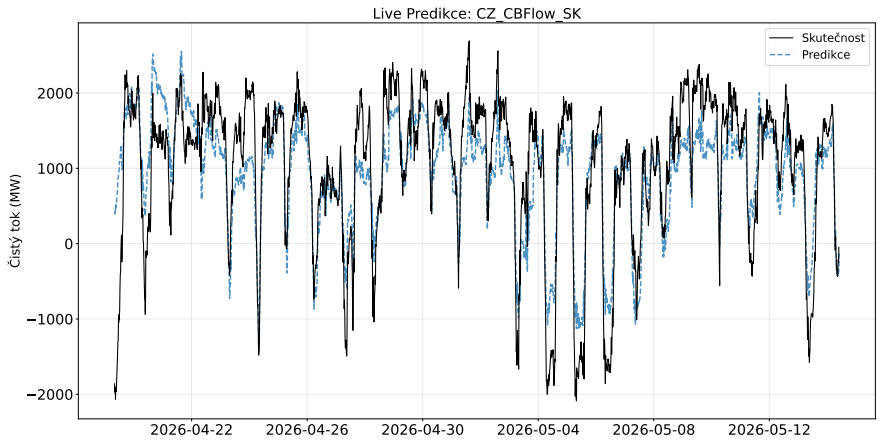
Obrázek 4.10: Celkový průběh reálných predikcí na profilu CZ → PL



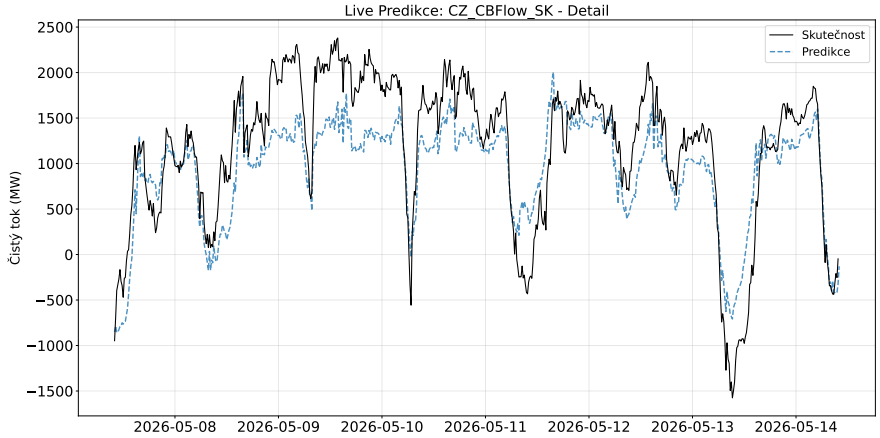
Obrázek 4.11: Detailní pohled na predikci na profilu CZ → PL

4.2.6 Profil CZ → SK

Slovenský profil vykazuje MAE 413,90 MW. Vysoká korelace 0,8586 naznačuje, že model velmi přesně identifikuje časové okamžiky změn trendu i při vysokých objemech tranzitu typických pro tuto hranici.



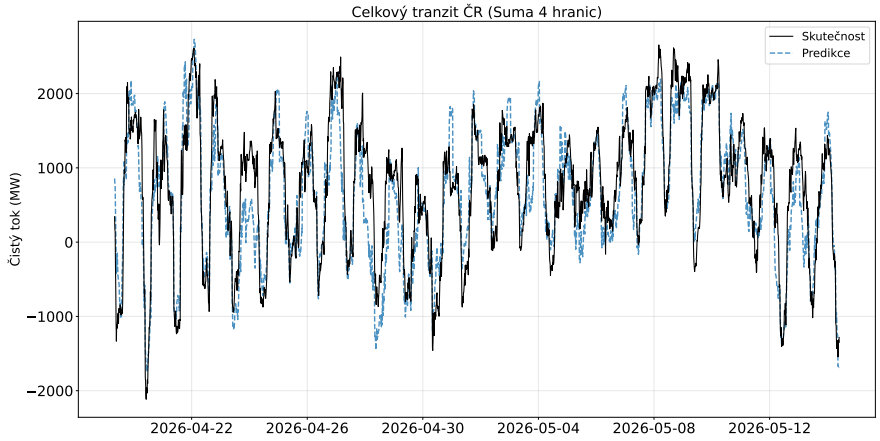
Obrázek 4.12: Celkový průběh reálných predikcí na profilu CZ → SK



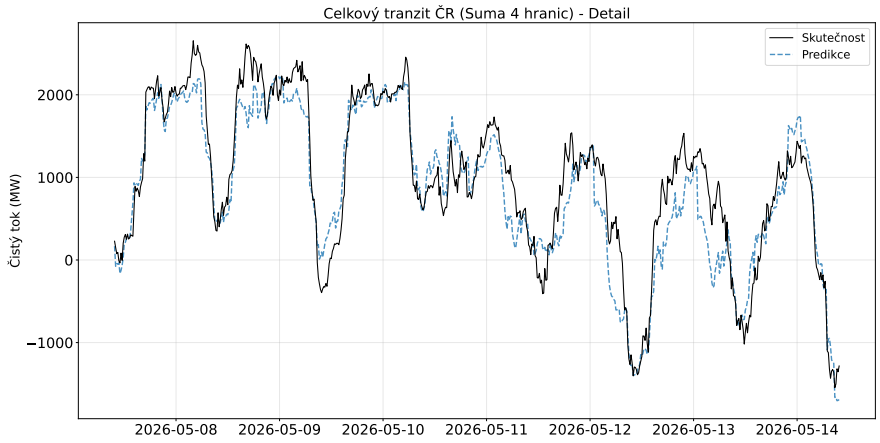
Obrázek 4.13: Detailní pohled na predikci na profilu CZ → SK

4.2.7 Celkový tranzit ČR

Nejlepších výsledků v rámci reálného provozu bylo dosaženo při predikci celkového tranzitu České republiky. Agregovaný model dosáhl koeficientu determinace $R^2 = 0,7613$ a nejvyšší korelace 0,8866 z celého testovaného souboru.



Obrázek 4.14: Celkový průběh reálných predikcí celkového tranzitu ČR



Obrázek 4.15: Detailní pohled na predikci celkového tranzitu ČR

Průměrná absolutní chyba MAE 332,84 MW je vzhledem k rozsahu sčítaných toků velmi příznivá. Vizuální analýza (Obrázky 4.14 a 4.15) potvrzuje, že model dokáže precizně sledovat globální trend bilance soustavy, což je klíčové pro celkové dispečerské plánování.

4.2.8 Shrnutí

Analýza reálného provozu potvrdila, že navržený systém je schopen autonomně stahovat data, provádět jejich harmonizaci a generovat predikce. Přestože metriky v reálném čase vykazují očekávaný pokles přesnosti oproti validačnímu vyhodnocení, modely si zachovaly vysokou korelaci se skutečným vývojem toků.

Nejstabilnějších výsledků bylo dosaženo u celkového tranzitu ČR ($R^2 = 0,7613$) a na slovenském profilu ($R^2 = 0,7099$). Získané poznatky naznačují, že systém je připraven pro praktické nasazení, přičemž dalším krokem k vyšší přesnosti by mohlo být zkrácení intervalu aktualizací predikcí.

Tabulka 4.2: Metriky modelů dosažené v reálném provozu (20. 4. – 14. 5. 2026)

Profil	**MAE [MW]**	**RMSE [MW]**	**MAPE [%]**	R^2 **[-]**	**Corr [-]**
CZ → AT	288,38	365,11	94,52	0,4747	0,7873
CZ → DE_LU	426,19	547,56	313,69	0,6821	0,8586
CZ → PL	268,09	346,11	91,52	0,6391	0,8098
CZ → SK	413,90	525,20	140,54	0,7099	0,8586
$\sum CZ$	332,84	436,83	91,28	0,7613	0,8866

4.3 Dopady kategorií příznaků na modely

Tato podkapitola se věnuje hloubkové analýze toho, jak jednotlivé kategorie vstupních dat skutečně přispívají k přesnosti modelů. Pro srovnání zde využíváme metriku průměrného marginálního přínosu, která odhaluje, o kolik se výkonnost modelu reálně zlepší (nebo zhorší) přidáním konkrétní skupiny příznaků při zachování ostatních vstupů.

Na rozdíl od prostého srovnání průměrných výsledků všech modelů s daným příznakem a bez něj, tento přístup eliminuje zkreslení způsobené synergií mezi různými skupinami dat. Metoda je založena na principu ablační studie a vychází z výpočtu Shapleyho hodnoty [Sha53; LMG80; Grö06], kdy jsou porovnávány dvojice modelů, které se liší právě v jedné sledované kategorii příznaků. Striktní dopad $I_{\text{strict}}(g)$ skupiny g je pak definován jako:

$$I_{\text{strict}}(g) = \frac{1}{N_g} \sum_{S \in \mathcal{P}_g} \left(R^2(S) - R^2(S \setminus \{g\}) \right) \tag{4.1}$$

Kde:

- $I_{\text{strict}}(g)$ je výsledný marginální dopad skupiny příznaků g,
- S je testovaná kombinace tříd příznaků, která obsahuje skupinu g (tedy $g \in S$),
- $S \setminus \{g\}$ je odpovídající kombinace tříd bez skupiny g,
- $\mathcal{P}_g$ je množina všech takových dvojic $(S, S \setminus \{g\})$, které byly v rámci experimentu obě evaluovány,
- N_g je celkový počet těchto nalezených dvojic ($N_g = |\mathcal{P}_g|$),
- $R^2(S)$ je hodnota metriky (koeficientu determinace) pro danou sadu.

Tento výpočet poskytuje robustní odhad reálné užitečnosti každého datového zdroje. Grafické srovnání těchto přínosů pro všechny analyzované hraniční profily je zobrazeno v Podkapitole 4.3.4. Výsledky umožňují identifikovat, které kategorie jsou pro predikci tranzitních toků klíčové a které naopak vnášejí do modelu šum.

4.3.1 Dominantní příznakové sady (C, N, E)

Analýza marginálního přínosu napříč všemi hraničními profily ukazuje jasnou dominanci energetických dat nad ostatními typy příznaků. Výsledky naznačují, že modely se při predikci fyzikálních toků opírají primárně o stav soustavy jak v cílové zemi, tak v zemích sousedních:

- **Skupiny příznaků C a N (Lokální a sousední energetika):** Tyto kategorie představují dominantní příznaky pro všechny sledované profily, přičemž jejich marginální přínos se stabilně pohybuje v nejvyšších hodnotách. Například na profilu CZ → DE_LU (Obrázek 4.17) dosahují hodnoty u třídy C 0,2299 a 0,2157 u třídy N. Podobně u profilu CZ → SK (Obrázek 4.19) je přínos třídy C 0,1817 a třídy N 0,1775.
- **Skupina příznaků E (Ostatní sousedící země):** Význam této kategorie vykazuje největší rozdíly mezi jednotlivými hranicemi. Zatímco u profilů CZ → DE_LU (0,1664), CZ → SK (0,1445) a CZ → PL (0,1376) hraje třída E klíčovou roli, u profilu CZ → AT je její přínos výrazně nižší a dosahuje pouze hodnoty 0,0432. Tento rozdíl naznačuje, že tok na rakouské hranici je více izolován od širšího evropského tranzitu než v případě ostatních sousedů.

4.3.2 Význam autoregrese a počasí (A, B, W)

Skupiny A (Autoregresivní hodnoty), B (Baseline) a W (Počasí) vykazují stabilní, avšak výrazně nižší přínos než přímá energetická data:

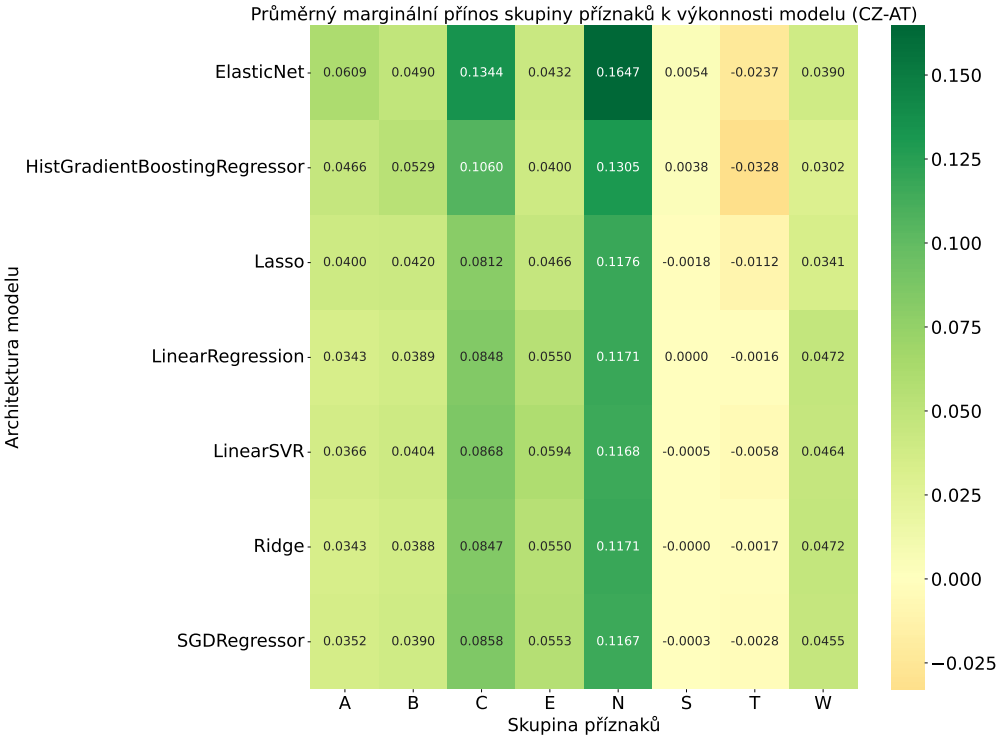
- **Autoregrese (A, B):** Přínos těchto tříd se pohybuje typicky v rozmezí 0,01 až 0,06. Například u profilu CZ → AT (Obrázek 4.16) má třída A pro `ElasticNet` přínos 0,0609.
- **Meteorologická data (W):** Přínos se pohybuje mezi 0,01 až 0,04. Nižší hodnota je pravděpodobně způsobena tím, že vliv počasí je již nepřímo obsažen v predikcích výroby z obnovitelných zdrojů, které jsou součástí energetických tříd C, N a E.

4.3.3 Negativní a zanedbatelný vliv (S, T)

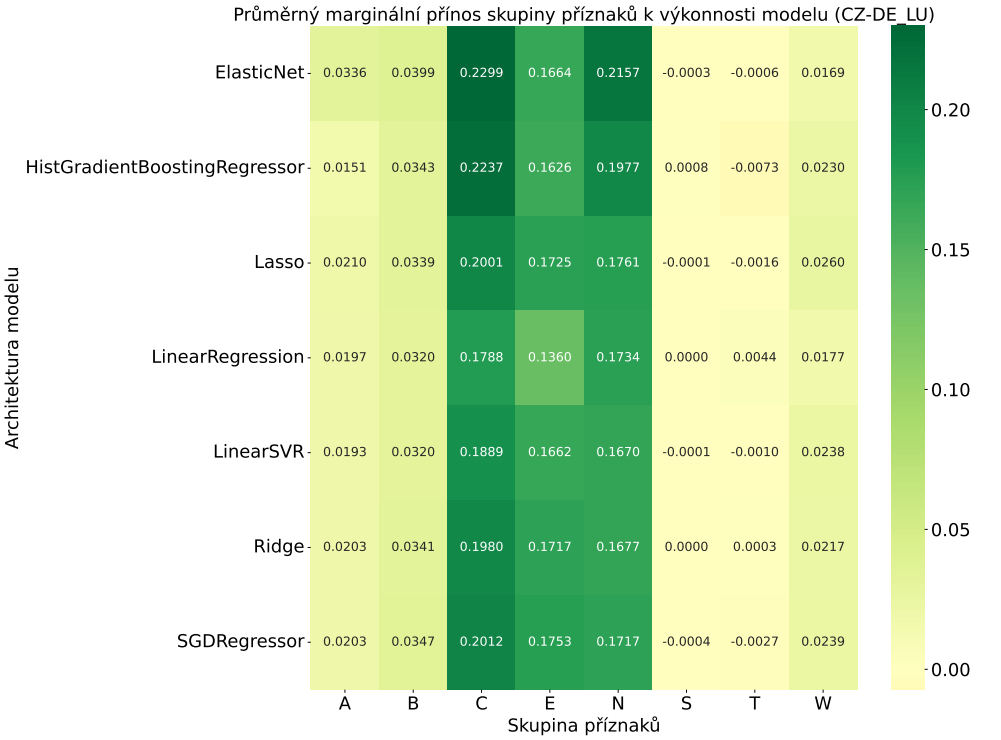
Analýza odhalila, že některé skupiny příznaků mohou výkonnost modelu mírně snižovat, což poukazuje na možné zavádění nerelevantních příznaků při trénování:

- **Časové příznaky (T):** Tato skupina vykazuje u mnoha modelů negativní marginální přínos, například −0,0328 u `HistGradientBoostingRegressor` na profilu CZ → AT (Obrázek 4.16) nebo −0,0190 u stejného modelu na profilu CZ → PL (Obrázek 4.18).
- **Speciální hodnoty (S):** Cenové diferenciály mají téměř nulový vliv na všech profilech. Například u profilu CZ → DE_LU se přínos třídy S pohybuje od −0,0004 po 0,0008 (Obrázek 4.17). Tento výsledek lze vysvětlit tím, že cenové diferenciály jsou již nepřímo zahrnuté v plánovaných obchodních výměnách, které jsou součástí energetických dat. Modely tak tuto informaci získávají z jiných příznaků a třída S se stává redundantní.

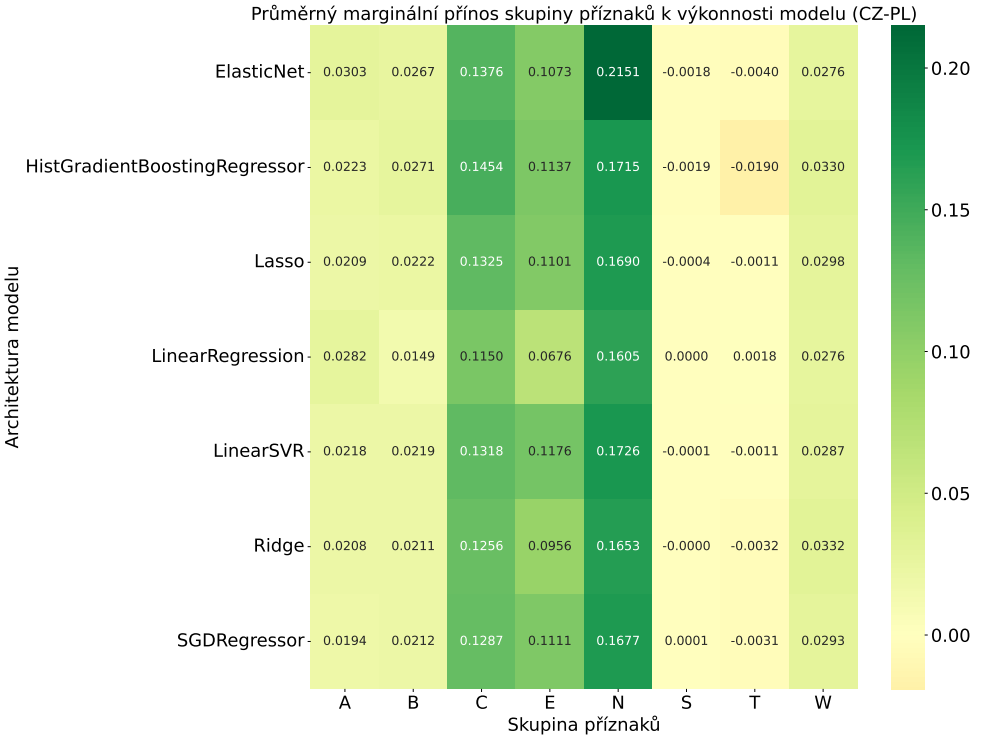
4.3.4 Srovnání v grafu



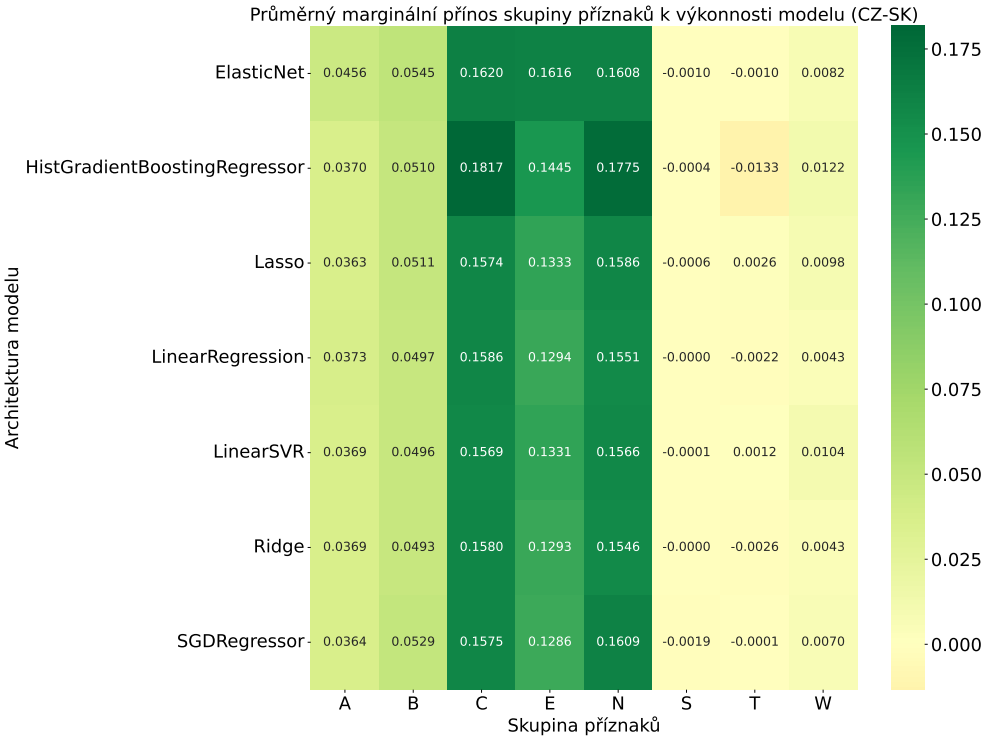
Obrázek 4.16: Marginální přínos R^2 na profilu CZ → AT



Obrázek 4.17: Marginální přínos R^2 na profilu CZ → DE_LU



Obrázek 4.18: Marginální přínos R^2 na profilu CZ → PL



Obrázek 4.19: Marginální přínos R^2 na profilu CZ → SK

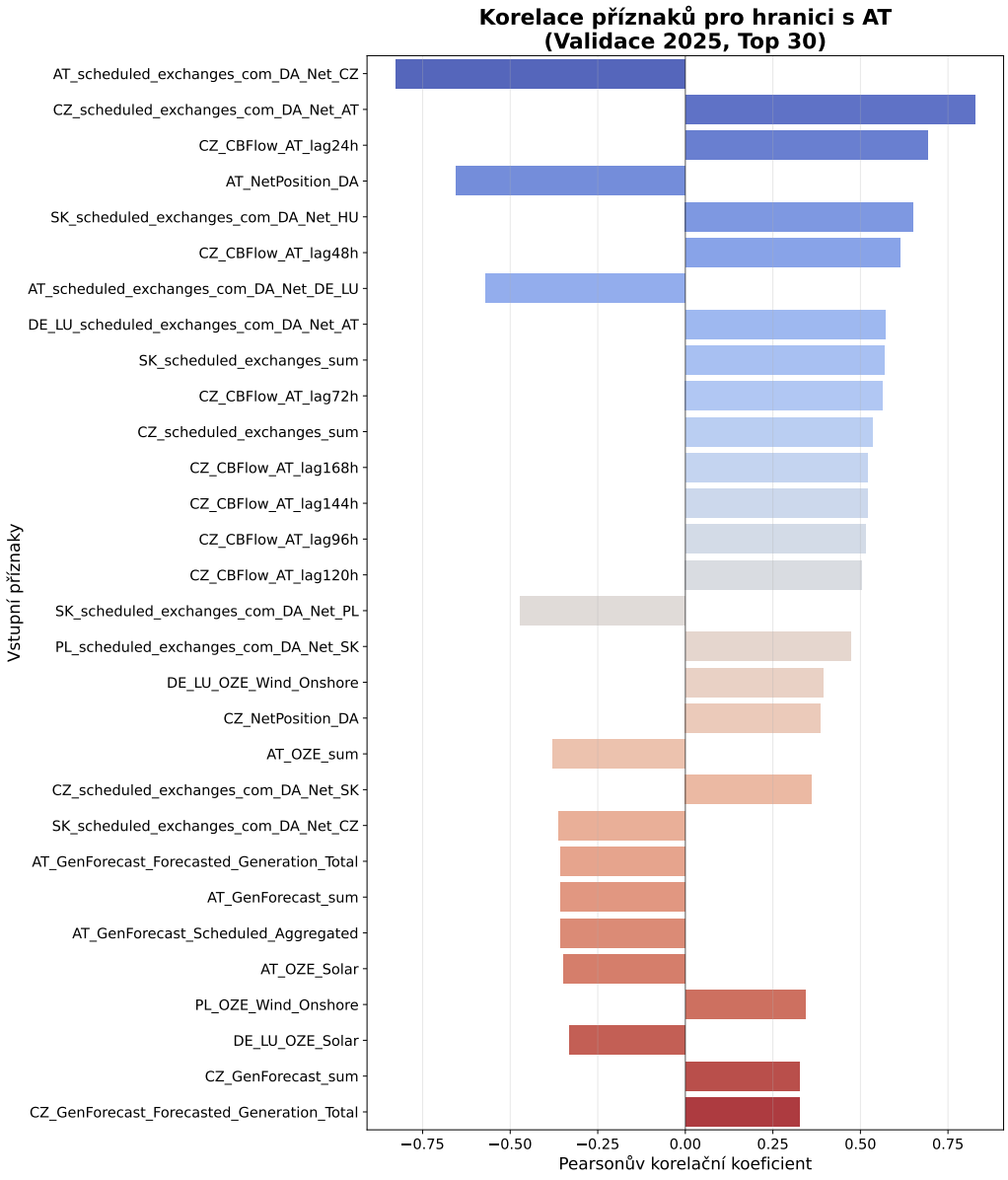
4.4 Korelační analýza příznaků

V této kapitole je provedena podrobná analýza korelací mezi cílovými veličinami a jednotlivými příznaky ze všech sledovaných kategorií. Hodnoty byly vypočteny pomocí vzorce Pearsonova korelačního koeficientu (vzorec 2.5) v Kapitole 2.6.1. Korelační koeficienty poskytují zásadní vhled do fyzikálních a ekonomických vazeb v rámci evropské energetické soustavy a slouží jako kvantitativní podklad pro výběr jednotlivých příznaků v datových sadách.

Analýza ukazuje, že nejdůležitějším příznakem pro všechny hraniční profily jsou plánované obchodní výměny. Významnou roli však hrají i meteorologická data a produkce z obnovitelných zdrojů v sousedních zemích.

4.4.1 Profil CZ → AT

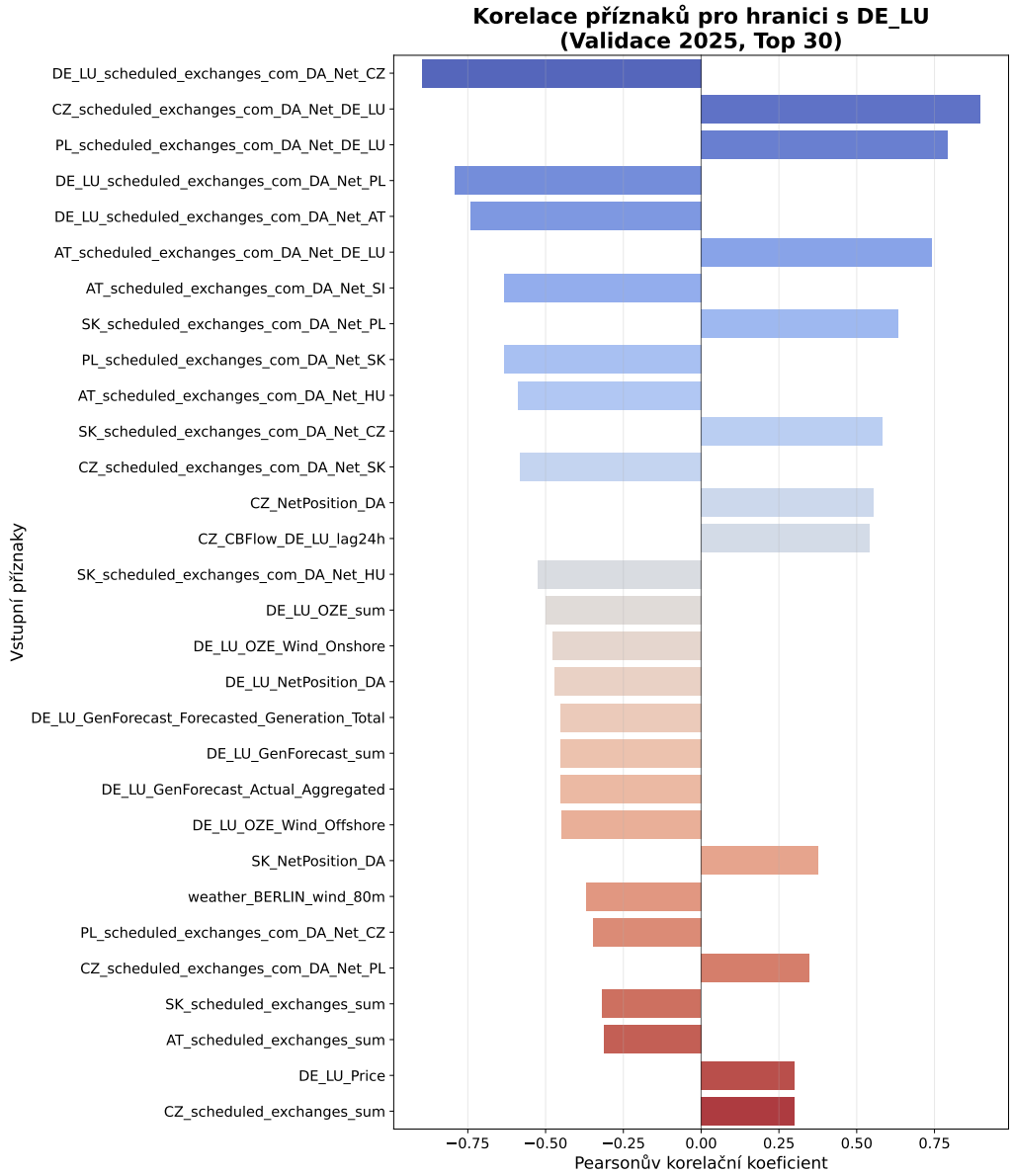
U rakouského profilu byla zjištěna nejsilnější vazba na přímou obchodní výměnu mezi ČR a Rakouskem s hodnotou korelace 0,7650. Specifikem tohoto profilu je relativně vysoká citlivost na teplotní podmínky, reprezentovaná příznaky *hdh* (Heating Degree Hours) pro Vídeň a Bratislavu s korelací přibližně 0,55. To naznačuje, že toky na této hranici jsou silně ovlivněny potřebou vytápění v regionu střední Evropy.



Obrázek 4.20: Korelační mapa příznaků pro profil CZ → AT

4.4.2 Profil CZ → DE_LU

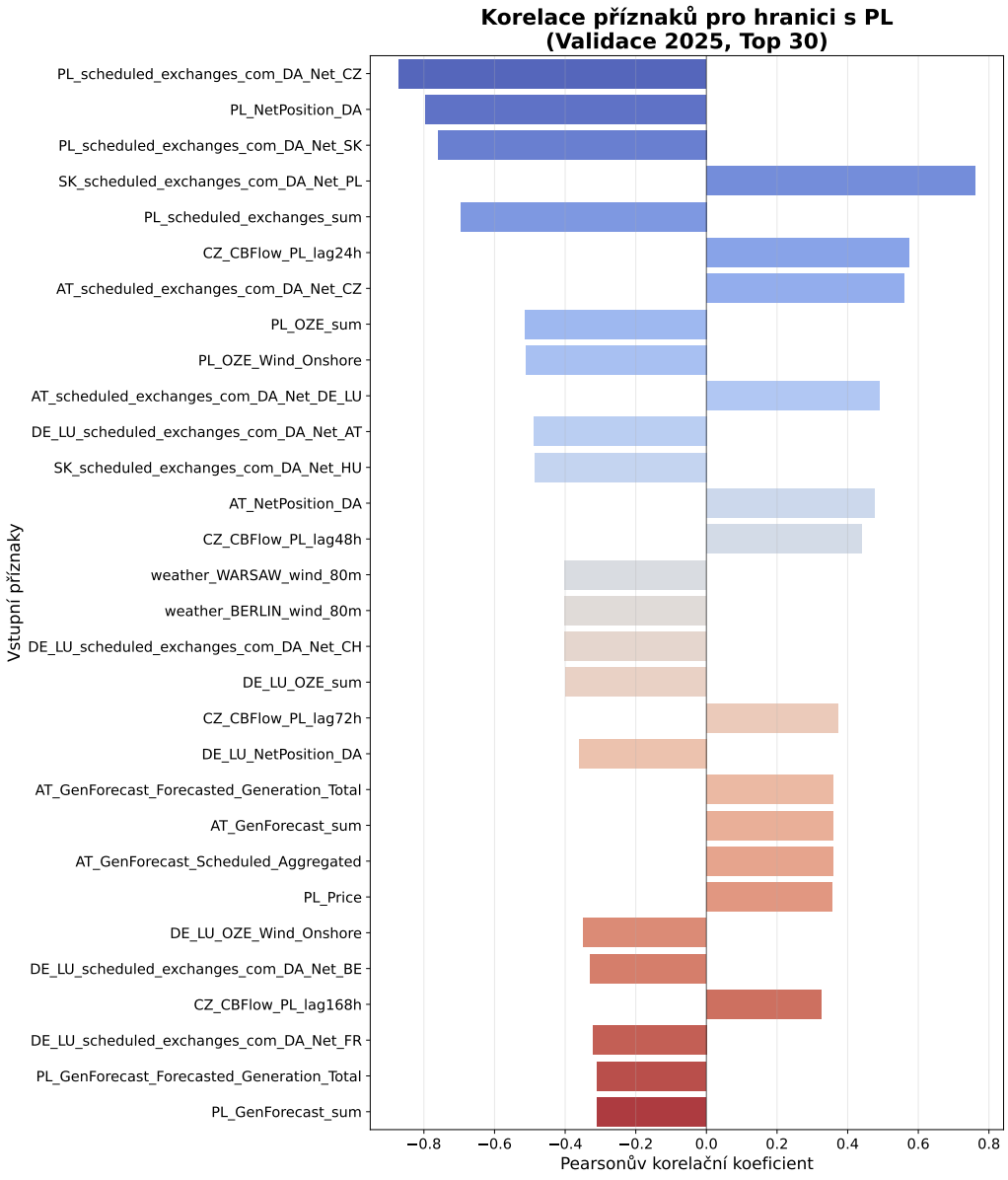
Profil s Německem vykazuje komplexní závislost na celoevropské situaci. Kromě přímé obchodní výměny (0,8574) je patrná silná záporná korelace s obchodními toky mezi Německem a Polskem (−0,7742). Zásadním zjištěním je potvrzený vliv německé větrné energetiky – korelace s pevninskými větrnými elektrárnami (Wind Onshore) v Německu dosahuje hodnoty −0,5417, u mořských (Offshore) pak −0,4826. Záporné znaménko indikuje, že při vysoké výrobě v Německu dochází k masivnímu tranzitu směrem do ČR.



Obrázek 4.21: Korelační mapa příznaků pro profil CZ → DE_LU

4.4.3 Profil CZ → PL

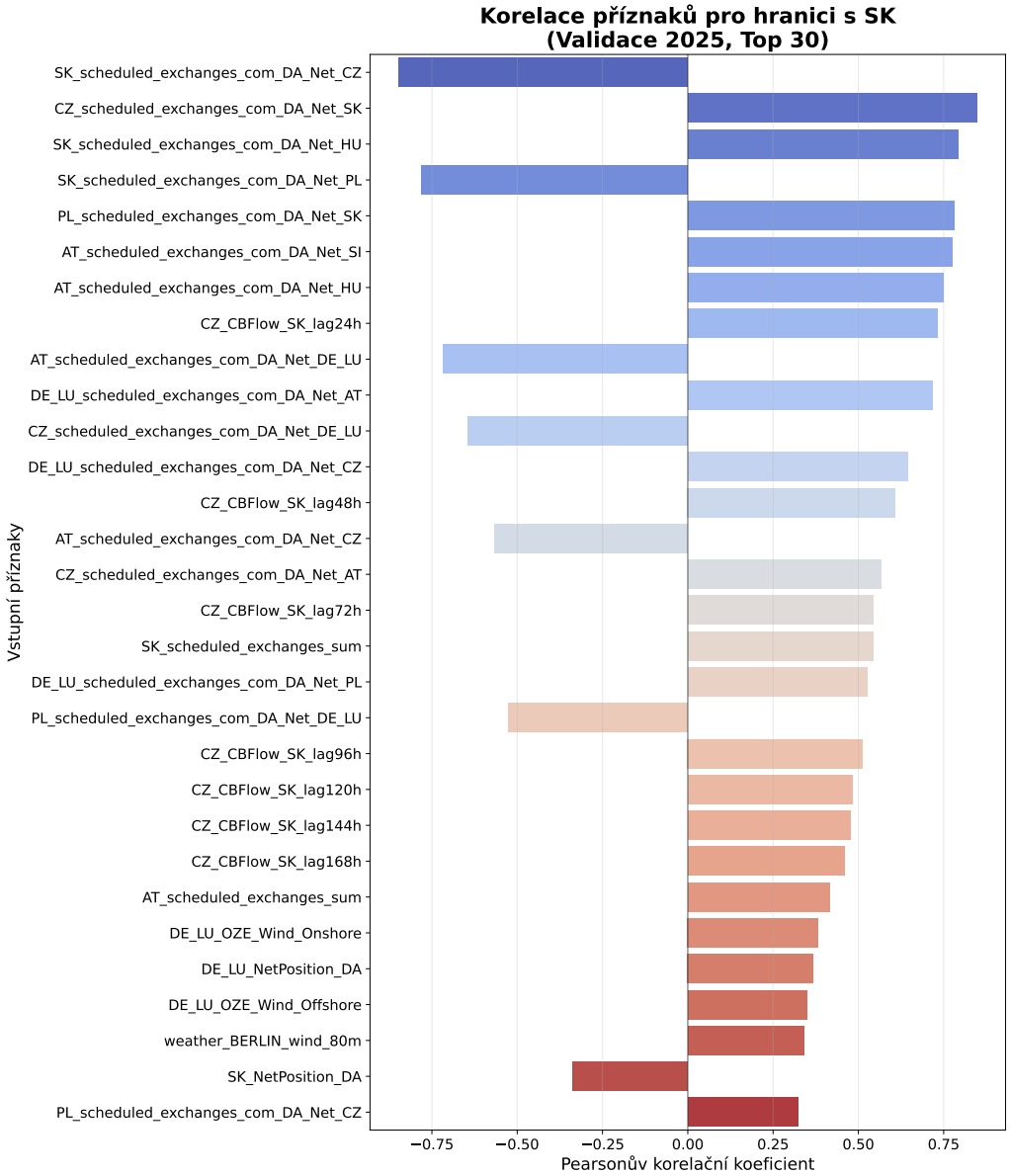
Na polské hranici dominuje závislost na plánované výměně směrem do ČR (−0,8498) a celkové obchodní pozici Polska (−0,7767). Významným faktorem je rovněž polská větrná výroba s korelací −0,5214. Fyzikální toky na tomto profilu jsou tak silně ovlivněny přetoky energie směřujícími z německých a polských větrných parků dále na jih.



Obrázek 4.22: Korelační mapa příznaků pro profil CZ → PL

4.4.4 Profil CZ → SK

Slovenský profil vykazuje nejsilnější provázanost s obchodními plány směrem do Maďarska, konkrétně na relacích SK → HU (0,8200) a AT → HU (0,7961). Tato zjištění kvantitativně potvrzují roli České republiky jako klíčového tranzitního uzlu pro energii směřující z trhů severní a střední Evropy dále na jihovýchod kontinentu.



Obrázek 4.23: Korelační mapa příznaků pro profil CZ → SK

4.4.5 Shrnutí korelační analýzy

Výsledky provedené korelační analýzy lze shrnout do následujících klíčových bodů:

- **Dominance obchodních plánů:** Plánované obchodní výměny (Scheduled Exchanges) jsou nejsilnějším příznakem u všech profilů, s nejvyšší korelací 0,8574 zjištěnou na německé hranici.
- **Vliv obnovitelných zdrojů:** Výroba z větrných elektráren v Německu a Polsku vykazuje silnou zápornou korelaci s toky do ČR (až −0,54), což potvrzuje přímý fyzikální vliv nadvýroby v těchto zemích na vznik neplánovaných tranzitních toků přes české území.
- **Teplotní citlivost:** Potřeba vytápění v regionu střední Evropy (příznaky hdh) má největší dopad na rakouský profil, kde korelace dosahuje hodnoty přibližně 0,55.
- **Potvrzení tranzitního charakteru:** Silná provázanost toků na slovenské hranici s obchodními plány směřujícími do Maďarska (korelace 0,8200) kvantitativně dokládá roli ČR jako klíčového tranzitního uzlu v severojižním směru.

Závěr a shrnutí výsledků

5

Hlavním cílem této bakalářské práce bylo navrhnout a implementovat výpočetní systém pro predikci fyzických tranzitních toků elektrické energie přes hranice České republiky s využitím metod strojového učení a volně dostupných datových zdrojů. Vzhledem k dynamické transformaci evropské energetiky a rostoucímu podílu obnovitelných zdrojů energie se jedná o vysoce aktuální a komplexní problém, který má přímý dopad na stabilitu, bezpečnost a ekonomickou efektivitu přenosové soustavy.

5.1 Shrnutí výsledků

V rámci práce byl úspěšně vytvořen robustní softwarový nástroj v jazyce Python, který dokáže plně automatizovaně stahovat, čistit a sjednocovat data z platformy ENTSO-E Transparency a meteorologického API Open-Meteo. Byla navržena komplexní pipeline pro předzpracování dat, která řeší sjednocení časového rozlišení, prostorové průměrování vývoje počasí i generování odvozených a zpožděných příznaků.

Experimentální část práce vyhodnotila několik modelů strojového učení (např. `ElasticNet`, `LinearSVR`, `Lasso`, `SGDRegressor`) na rozsáhlých sadách příznaků. Výsledky prokázaly, že využití agregovaných energetických a meteorologických dat vede k masivnímu zlepšení přesnosti predikce oproti základním modelům. Implementovaný systém byl navíc úspěšně otestován v reálném provozu po dobu více než tří týdnů, čímž byla ověřena jeho schopnost fungovat jako automatizovaný prediktivní nástroj.

5.2 Odpovědi na výzkumné otázky z Kapitoly 1.3

Na základě provedených experimentů a korelační analýzy lze formulovat následující odpovědi na výzkumné otázky z úvodu práce:

1. **Jaké faktory nejvýznamněji ovlivňují tranzitní toky?**
   Nejsilnějším příznakem se ukázaly být plánované obchodní výměny mezi jednotlivými zeměmi. Zásadní vliv má rovněž produkce z obnovitelných zdrojů v sousedních zemích – zejména vysoká výroba z větrných elektráren v Německu a Polsku vykazuje silnou korelaci s toky do ČR, což vede k neplánovaným přetokům na naše území. U některých hranic (např. s Rakouskem) byl prokázán i vliv teplotních podmínek v souvislosti s potřebou vytápění.

2. **Lze nahradit komplexní fyzikální modely datovými modely?**
   Výsledky práce ukazují, že datové modely představují funkční alternativu. Bez jakékoliv znalosti vnitřní topologie sítě nebo fyzikálních parametrů vedení dokázaly modely strojového učení na základě historických dat vysvětlit téměř 96 % rozptylu ($R^2 = 0,9589$) u celkového tranzitu na historických datech a zachovat korelaci přes 0,88 v reálném provozu. Fyzikální modely tak lze pro účely krátkodobé predikce úspěšně aproximovat.

3. **Jaký vliv má zahrnutí počasí do modelu?**
   Zahrnutí meteorologických dat přináší modelu stabilitu a zlepšuje predikci, avšak analýza marginálního přínosu ukázala, že jeho striktní dopad je nižší než u přímých energetických dat. Hlavním důvodem je skutečnost, že velká část informací o počasí je již nepřímo dostupná v predikcích výroby ze solárních a větrných elektráren. Vliv počasí je tak pro model užitečným doplňkem, ale není zcela nezbytným primárním zdrojem.

5.3 Zhodnocení predikce celkového tranzitu a budoucí rozvoj

Nejlepších a nejstabilnějších výsledků bylo dosaženo při predikci celkové tranzitní bilance České republiky. Agregací predikcí jednotlivých přeshraničních profilů dochází ke vzájemné kompenzaci lokálních chyb, což vedlo k dosažení průměrné absolutní chyby pouhých 173,52 MW na validačních datech a 332,84 MW při nasazení v reálném čase. Vzhledem k tomu, že se fyzikální toky pohybují v řádech nízkých tisíců megawattů, jde o velmi nadějný výsledek s potenciálem pro praktické dispečerské řízení. Pokles přesnosti při živém testování zároveň věrně ilustruje náročnost reálného nasazení, kde se modely musí potýkat s dynamicky se měnícím energetickým mixem a neustálým připojováním nových OZE, které historická data neobsahují.

Možnosti budoucího rozvoje:

Na tuto práci lze navázat v několika směrech. Prvním možným vylepšením je zkrácení intervalu pro stahování dat a generování predikcí z dvanáctihodinového cyklu na hodinový, což by umožnilo dispečerům pružněji reagovat na nečekané změny v počasí či výpadky v síti. Dále by bylo vhodné rozšířit trénovací datasety o delší historické období, aby modely lépe zachytily dlouhodobé jevy v síti. S narůstajícím objemem dostupných dat by se také otevřel prostor pro aplikaci komplexnějších architektur hlubokého učení, jako jsou rekurentní neuronové sítě či modely založené na architektuře Transformer, které by mohly vykazovat lepší výsledky.

Bibliografie

[Bol13] BOLDIŠ, Zbyněk. Czech electricity grid challenged by German wind. *Europhysics News*. 2013, roč. 44, č. 4, s. 16–18. Dostupné z DOI: 10.1051/epn/2013401.

[ČEP25] ČEPS, A.S. *Data do kapsy 2025*. Praha: ČEPS, a.s., 2025. Dostupné také z: https://www.ceps.cz/cs/materialy-ke-stazeni. [online; cit. 2026-05-14].

[Ene26] ENERGIEID. *entsoe-py: Python client for the ENTSO-E Transparency Platform API* [https://github.com/EnergieID/entsoe-py]. GitHub, 2026.

[ENT26] ENTSO-E. *ENTSO-E Transparency Platform* [online]. 2026. [cit. 2026-04-30]. Dostupné z: https://transparency.entsoe.eu/.

[Evr13] EVROPSKÁ KOMISE. *Nařízení Komise (EU) č. 543/2013 ze dne 14. června 2013 o předkládání a zveřejňování údajů na trzích s elektřinou a o změně přílohy I nařízení Evropského parlamentu a Rady (ES) č. 714/2009* [http://data.europa.eu/eli/reg/2013/543/oj]. 2013. Dostupné online.

[Grö06] GRÖMPING, Ulrike. Relative Importance for Linear Regression in R: The Package relaimpo. *Journal of Statistical Software*. 2006, roč. 17, č. 1, s. 1–27. Dostupné z DOI: 10.18637/jss.v017.i01.

[Hel18] HELBICH, Jiří. *Dispečerské řízení* [online]. ČEPS, a.s., 2018. [cit. 2026-05-11]. Dostupné z: http://home.zcu.cz/~nohac/E2/Dispecerske_rizeni_CEPS_ZCU-ProStudenty.pdf. Prezentace pro pedagogické účely.

[LMG80] LINDEMAN, Richard Harold; MERENDA, Peter F.; GOLD, Ruth Z. *Introduction to Bivariate and Multivariate Analysis*. Glenview, IL: Scott, Foresman, 1980.

[Ped+26] PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V. et al. *Choosing the right estimator — scikit-learn documentation* [https://scikit-learn.org/stable/machine_learning_map.html]. 2026. Accessed: 2026-04-30.

[Pet26] PETROVICH, Beatrice. *European Electricity Review 2026*. 2026-01. Report. Ember. Dostupné také z: https://ember-energy.org/app/uploads/2026/01/EMBER-Report-European-Electricity-Review-2026.pdf. Zveřejněno 22. ledna 2026.

[Sha53] SHAPLEY, Lloyd S. A Value for n-Person Games. In: KUHN, Harold William; TUCKER, Albert William (ed.). *Contributions to the Theory of Games (AM-28), Volume II*. Princeton: Princeton University Press, 1953, s. 307–318.

[Zip23] ZIPPENFENIG, Patrick. *Open-Meteo.com Weather API*. 2023. Dostupné z DOI: 10.5281/zenodo.7970649.

Seznam obrázků

2.1 Roční toky energie přes hranice ČR za rok 2025. Zdroj: [ČEP25] 12
2.2 Ukázka sinové a kosinové transformace na časových příznacích 15
2.3 Ukázka sinové a kosinové transformace na časových příznacích 16
2.4 Ukázka binárních a dalších časových příznaků 16
2.5 Vzorec použitý pro průměrování 19
2.6 Rozhodovací proces při výběru modelu podle scikit-learn. Zdroj: [Ped+26] 20

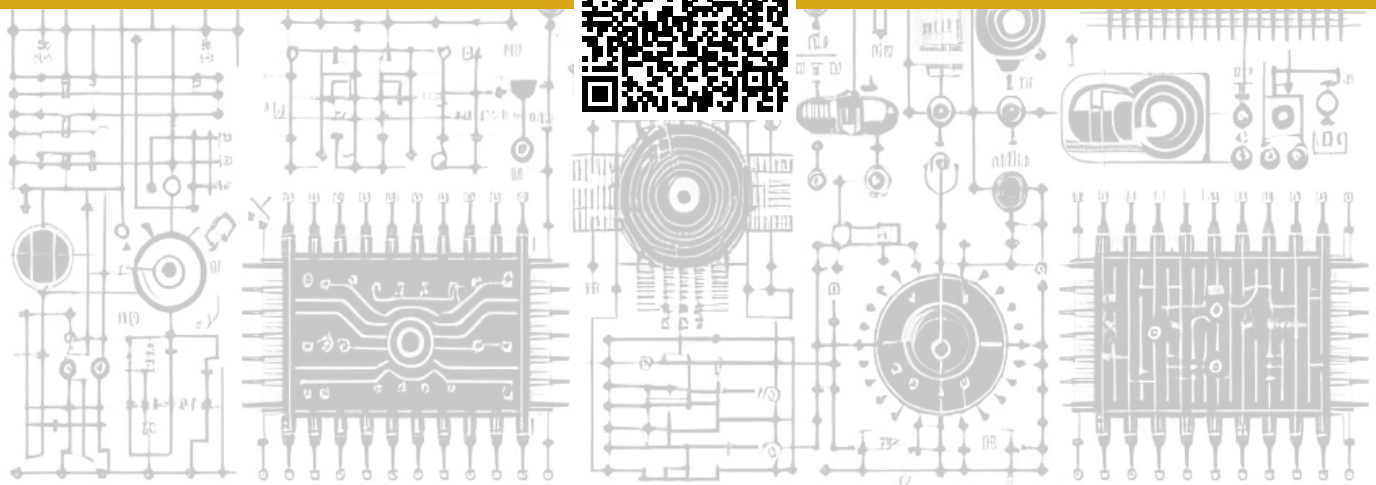
3.1 Diagram projektu predikce tranzitu 26

4.1 Srovnání predikcí s reálnými toky na profilu CZ → AT 35
4.2 Srovnání predikcí s reálnými toky na profilu CZ → DE_LU 36
4.3 Srovnání predikcí s reálnými toky na profilu CZ → PL 37
4.4 Srovnání predikcí s reálnými toky na profilu CZ → SK 38
4.5 Srovnání predikcí celkového tranzitu ČR 39
4.6 Celkový průběh reálných predikcí na profilu CZ → AT 41
4.7 Detailní pohled na predikci na profilu CZ → AT 41
4.8 Celkový průběh reálných predikcí na profilu CZ → DE_LU 42
4.9 Detailní pohled na predikci na profilu CZ → DE_LU 42
4.10 Celkový průběh reálných predikcí na profilu CZ → PL 43
4.11 Detailní pohled na predikci na profilu CZ → PL 43
4.12 Celkový průběh reálných predikcí na profilu CZ → SK 44
4.13 Detailní pohled na predikci na profilu CZ → SK 44
4.14 Celkový průběh reálných predikcí celkového tranzitu ČR 45
4.15 Detailní pohled na predikci celkového tranzitu ČR 45
4.16 Marginální přínos R^2 na profilu CZ → AT 50
4.17 Marginální přínos R^2 na profilu CZ → DE_LU 50
4.18 Marginální přínos R^2 na profilu CZ → PL 51
4.19 Marginální přínos R^2 na profilu CZ → SK 51
4.20 Korelační mapa příznaků pro profil CZ → AT 53
4.21 Korelační mapa příznaků pro profil CZ → DE_LU 54
4.22 Korelační mapa příznaků pro profil CZ → PL 55

4.23 Korelační mapa příznaků pro profil CZ → SK 56

Seznam tabulek

2.1 Souhrnný přehled dostupnosti a technických parametrů datových zdrojů. 17

4.1 Souhrnné porovnání metrik pro jednotlivé hraniční profily a celkový tranzit ČR 39

4.2 Metriky modelů dosažené v reálném provozu (20. 4. – 14. 5. 2026) . . . 46