

Предикција производње и потрошње електричне енергије на основу временских услова и добра дана

Милош Нешковић Е2 91/2025, Ненад Дубовац Е2 98/2025

1. Дефиниција проблема

Анализом скупа података потребно је провјерити у којој мјери временски услови и доба дана утичу на производњу електричне енергије која долази из обновљивих извора. Поред тога потребно је извршити анализу у којој мјери потражња за електричном енергијом зависи од истих параметара.

2. Мотивација проблема рјешаваног у пројекту

Детаљном анализом податка могао би се креирати модел који би предводио да ли ће бити довољно електричне енергије која је створена из обновљивих извора сходно потребама за тај временски период узимајући у обзир и временске услове како би се на оптималан начин могло извршити оптерећење мреже.

3. Релевантна литература

3.1. A comparative study of Forcasting electricity consumption using machine learning models [[линк](#)]

Тема рада: У овом раду коришћено је више модела за предикцију производње и потрошње електричне енергије и тражен је најбољи у циљу оптималне производње енергије у складу са потребама.

Подаци: Коришћени су мјесечни подаци о потрошњи електричне енергије за седам земаља (Норвешка, Швајцарска, Малезија, Египат, Алжир, Бугарска и Кенија) за период од 2007. до 2016. године. Подаци о потрошњи у периоду од 2007. до 2015. године коришћени су као тренинг скуп, док су подаци за 2016. годину коришћени као тестни скуп податак.

Методологија: У раду је коришћено четири различита модела машинског учења за предикцију потрошње електричне енергије:

- ANN (Artificial Neural Network),

- ANFIS (Adaptive Neuro-Fuzzy Inference System),
- LSSVM (Least Squares Support Vector Machines),
- FTS (Fuzzy Time Series)

Сви модели су обучени и тестирали на истим подацима о потрошњи електричне енергије за сваку од седам земаља.

Евалуација рјешења: Коришћене су три различите метрике за евалуацију root mean square error (RMSE), average forecasting error, (AFE) performance parameter (PP). Од четири модела која су коришћени у овом раду FTS се показао као најбољи за дугорочно прогнозирање.

Закључак: Овај рад нам је релевантан јер се бави сличном тематиком и користи неуронску мрежу коју ћемо и ми користити у нашем раду.

3.2. Forecasting Short-Term Photovoltaic Energy Production to Optimize Self-Consumption in Home Systems Based on Real-World Meteorological Data and Machine Learning [[линк](#)]

Тема рада: Рад се бави изградњом модела за предвиђање производње електричне енергије са соларних панела коришћењем стварних метеоролошких података. Тражен је најбољи модел ради што прецизније процене производње електричне енергије ради оптимизације потрошње у кућним системима.

Подаци: Коришћени су подаци који су прикупљани са кровне фотонапонске инсталације. Подаци су прикупљани у интервалима од по 5 минута и агрегирани на 1 сат. Прикупљање података је трајало више од 2 године, сходно томе модели су тренирани на узорцима од приближно 18.000 сатних редова. Тренинг и тестни скуп су подељени у односу 80/20, где се ранији подаци користе за тренирање модела а каснији за тестирање.

Методологија: У раду су коришћена три модела машинског учења за предикцију производње електричне енергије:

- Linear Regression
- Random Forest
- XGBoost

Евалуација рјешења: Коришћене су четири метрике за евалуацију: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2).

Закључак: У раду планирамо да користимо Random Forest ili XGBoost алгоритам јер се показало да дају добре резултате у предвиђању производње електричне енергије, што је такође тема нашег пројекта.

3.3. Deep Learning for Multi-Output Regression Using Gradient Boosting [[Линк](#)]

Тема рада: Рад представља метод за рјешавање проблема регресије са више излаза. Тестира више модела неуронских мрежа, дискутује и тражи најбољи.

Подаци: Скуп података који је коришћен у овом раду представља комбинацију 17 скупова података. Ти скупови података припадају различитим доменима и имају варијације у броју инстанци, атрибута и циљних промјењивих.

Методологија: У овом раду примјењени су различити модели неуронских мрежа (NN, DNN, GBNN, GB-DNNR) које је аутор упоређивао са циљем проналаска најбољег.

Евалуација рјешења: Као метрике за евалуацију коришћени су MAE и RMSE.

Закључак: Овај рад је релевантан за наш пројекат јер показује да је могуће конструисати један модел који ефикасно рјешава проблем регресије са више излаза. Наш циљ је да покушамо креирати један модел који ће радити предикцију и потрошње и произвидње електричне енергије.

4. Скуп података

Податке смо пронашли на Kaggle сајту. Подаци се састоје из два скупа података. Први скуп података [[Линк](#)] представља податке везане за потрошњу и потражњу електричне енергије у периоду од 01.01.2015. до 31.12.2018. у периодима од сат времена и састоји се од 30064 реда и 29 колона. Од тих колона нама су релевантне time, generation solar, generation wind onshore, total load actual.

Други скуп података [[Линк](#)] садржи информације о временским условима у 5 највећих шпанских градова у истом временском периоду. Састоји се од 17 колона од којих су за наше истраживање најрелевантније dt_iso, temp, humidity, wind_speed, rain_1h, weather_description.

Од ова два одвојена скупа података направићемо један са релевантим колонама. Спајање ћемо вршити по времену.

5. Методологија

Прво је потребно извршити пондерисање података о временским условима на основу близине града неким од највећих паркова за производњу електричне енергије. Након тога потребно је спојити скупове података по заједничкој колони (вријеме). Креираћемо по 2 модела (Random Forest и XGBoost) за предкицију производње електричне енергије соларних панела, предикцију производње електричне енергије настале радом вјетроелектрана и на крају предикцију потрошње електирчне енергије.

На крају ћемо креирати и један multi-output модел и проверити да ли ће се он показати боље или лошије у односу на специјализоване моделе.

6. Метод евалуације

За одступање предиктоване вриједност од стварне користићемо RMSE или MAE метрику. Податке ћемо подјелити на тренинг, валидациони и тест скуп у односу 70:15:15.