

پیشبینی گاف نواری پروسکایت‌ها با استفاده از یادگیری ماشین

علیرضا صادقی^۱، محسن وفائی حسین آبادی^{۲*}

۱ - دانشجوی کارشناسی ارشد، شیمی، شیمی فیزیک، دانشکده علوم پایه، دانشگاه تربیت مدرس

sadeghi_alireza@modares.ac.ir

۲ - استادیار، شیمی، شیمی فیزیک، دانشکده علوم پایه، دانشگاه تربیت مدرس

m.vafaei@modares.ac.ir

واژگان کلیدی: پروسکایت، گاف نواری، یادگیری ماشین

۱- مقدمه

مواد بر پایه پروسکایت توجه زیادی را در زمینه های مختلف به خود جذب کرده است. دلایل این توجهات، تنوع ترکیبات، آسان بودن سنتز و خواص جذاب گوناگون این دسته از ترکیبات است [1,2]. گاف نواری یکی از خصوصیات تأثیرگذار این مواد برای کاربردهای مختلف، به ویژه در فوتوولتائیک، است. با این حال، هزینه و محدودیت سرعت محاسبه گاف نواری در آزمایشگاه چالش‌هایی را ایجاد کرده است. یکی از راه‌حل‌ها استفاده از روش‌های محاسباتی و مدل‌سازی با هزینه کم و سرعت بالا است.

یادگیری ماشین در سال‌های اخیر توسعه چشمگیری داشته و تأثیر بزرگی در علوم مختلف داشته است. از این رو هدف اصلی این پژوهش استفاده از مدل‌های یادگیری ماشین برای محاسبه گاف نواری پروسکایت‌ها است. از مزیت‌های استفاده از یادگیری ماشین در این پژوهش تفسیر پذیری محاسبات و امکان مهندسی خواص مواد در مطالعات بعدی است.

۲- روش

در ابتدا از مجموعه ۱۱۰۲۶ داده گاف نواری محاسبه شده با DFT توسط Nakajima و همکاران [3] استفاده شده است. سپس داده‌ها پردازش شده و داده‌های اضافی و خالی حذف شدند. سپس ویژگی‌های سایت‌های $X_B A$ پروسکایت‌ها مانند الکترون‌گاتیویته شعاع یونی انرژی یونیزاسیون، تعداد الکترون‌های ظرفیت و ... به عنوان مشخصه‌ها به داده‌ها اضافه شدند. سپس همبستگی بین مشخصه‌ها محاسبه و مشخصه‌هایی که ضرایب همبستگی همبستگی پیرسن بزرگتر از ۰.۸ داشتند حذف شدند. شکل (۱) همبستگی متغیرها را به صورت نقشه حرارتی نشان می‌دهد. پیش از فیت کردن مدل، داده‌ها به نسبت ۰.۸ و ۰.۲ برای آموزش و آزمایش داده تقسیم شدند. سپس مدل‌های یادگیری ماشین Random Forest، XGBoost و MLP برای محاسبه گاف نواری استفاده و خطای آنها محاسبه شد.

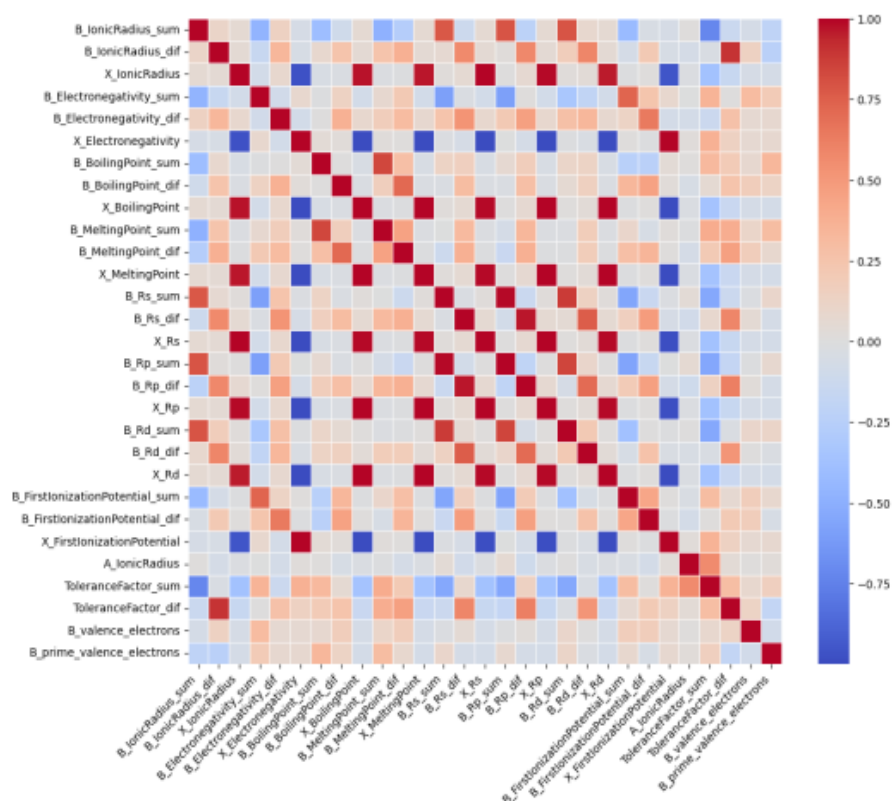
کدهای نوشته شده در این پژوهش به زبان برنامه نویسی پایتون و در محیط google colab نوشته و اجرا شدند و برای فیت کردن مدل‌های یادگیری ماشین از کتابخانه‌های scikit-learn و tensorflow استفاده شد.

۳- بحث و نتیجه گیری

از دو پارمتر MSE و R-Squared برای محاسبه دقت مدل‌های به کار رفته، استفاده شده است.

$$Mean Squared Error (MSE) = \left(\frac{1}{n}\right) * \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

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



شکل (۱): نقشه حرارتی مشخصه‌ها

محاسبه mse و R-Squared نشان داد به ترتیب مدل‌های XGBoost, Random Forest و MLP دقت بالاتری در محاسبه‌ی گاف نواری دارند. جدول (۱) خطاهای اندازه‌گیری مدل‌ها را با هم مقایسه کرده است.

نام مدل	R-Squared	MSE
XGBoost	0.75	0.22
Random Forest	0.74	0.23
MLP	0.71	0.26

جدول (۱): مقادیر R-Squared و MSE برای مدل‌های مختلف

یکی از ویژگی‌های مدل Random Forest، تفسیر پذیر بودن آن است [4]. به طوری که در آن وزن هر مشخصه در گاف نواری را میتوان استخراج کرد. این ویژگی میتواند کاربرد زیادی در زمینه طراحی و مهندسی خواص مواد داشته باشد. با استفاده از این مدل مشخص شد به ترتیب الکترون‌گاتیویته عنصر سایت B، تعداد الکترون‌های ظرفیت سایت B و نقطه جوش عنصر سایت B در حالت خالص بیشترین وزن را در گاف نواری پروسکایت‌ها دارند.

۴-منابع و مراجع

- [1] Oró-Solé, J. et al (2014). Synthesis, anion order and magnetic properties of $\text{RVO}_{3-x}\text{N}_x$ perovskites ($\text{R} = \text{La}, \text{Pr}, \text{Nd}; 0 \leq x \leq 1$). *J. Mater. Chem. C*, 2, 2212–2220
- [2] Shioyai, J. et al. (2020). Signature of band inversion in the perovskite thin-film alloys $\text{BaSn}_{1-x}\text{Pb}_x\text{O}_3$. *Phys. Rev. B* 101, 125125
- [3] Nakajima, T., & Sawada, K. (2017). Discovery of PB-Free perovskite solar cells via High-Throughput Simulation on the K Computer. *The Journal of Physical Chemistry Letters*, 8(19), 4826–4831
- [4] Robinson, R. L. M., Palczewska, A., Palczewski, J., & Kidley, N. (2017). Comparison of the predictive performance and interpretability of random forest and linear models on benchmark data sets. *Journal of Chemical Information and Modeling*, 57(8), 1773–1792