**Machine Learning with Wildfires**

1. **Di Giuseppe et al. (2016) “Predicting Fires Using ECMWF Weather Forecasts”**  
   Learned that coupling weather forecasts with fire risk indices significantly improves short-term wildfire prediction accuracy, especially in predicting ignition likelihood.
2. **Jain et al. (2020) “A Review of Machine Learning Applications in Wildfire Science”**  
   Provided a taxonomy of ML methods used in wildfire prediction and emphasized the importance of event-based rather than point-based modeling, which shaped our clustering approach.
3. **Veraverbeke et al. (2017) “Ecosystem Transitions and Wildfire Patterns in Boreal Forests”**  
   Highlighted how fuel moisture and vegetation type are strong determinants of fire duration, influencing our decision to include duff\_frac and cwd\_frac.
4. **Cortez & Morais (2007) “A Data Mining Approach to Predict Forest Fires Using Meteorological Data”**  
   Showed how decision trees and regression methods can be used effectively on tabular fire-weather data, validating our selection of tree-based baselines.
5. **Parisien et al. (2012) “Projected Wildfire Activity in Canada under Climate Change”**  
   Underlined the importance of long-term climate trends, which informed our feature engineering for lagged and seasonal weather variables.
6. **Yuan et al. (2022) “Deep Learning for Wildfire Spread Prediction”**  
   Demonstrated that ConvLSTM models capture spatiotemporal dependencies better than static models, reinforcing our plan to test deep learning architectures.
7. **Andela et al. (2019) “Global Trends in Fire Activity”**  
   Showed that human activity and land-use patterns strongly affect fire behavior, leading us to consider integrating land cover (covertype) into our predictors.

**Techniques & Methods Used**

1. **Ester et al. (1996) “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases (DBSCAN)”**  
   Provided the foundation for our spatiotemporal clustering of wildfire points into coherent events.
2. **Breiman (2001) “Random Forests”**  
   Explained the robustness of Random Forests to overfitting and missing data, supporting their use as a first baseline model.
3. **Chen & Guestrin (2016) “XGBoost: A Scalable Tree Boosting System”**  
   Influenced our choice to use XGBoost for its speed, scalability, and handling of tabular data with missing values.
4. **Ke et al. (2017) “LightGBM: A Highly Efficient Gradient Boosting Decision Tree”**  
   Taught us that histogram-based learning can handle large-scale datasets efficiently, relevant to our ~7M wildfire records.
5. **Lundberg & Lee (2017) “A Unified Approach to Interpreting Model Predictions (SHAP)”**  
   Inspired us to plan interpretability analysis for feature importance beyond built-in scores.
6. **Wang et al. (2021) “Permutation Importance: A Model-Agnostic Approach”**  
   Helped us validate our top features and reduce over-reliance on internal importance metrics.
7. **Seabold & Perktold (2010) “Statsmodels: Econometric and Statistical Modeling with Python”**  
   Guided our ANOVA and correlation analyses in the feature selection pipeline.

**Weather Data & Environmental Variables**

1. **Hijmans et al. (2005) “Very High Resolution Interpolated Climate Surfaces for Global Land Areas”**  
   Showed methods for aligning spatial climate grids with environmental datasets, informing our nearest-grid-point join.
2. **Abatzoglou (2013) “Development of Gridded Surface Meteorological Data for Ecological Applications”**  
   Highlighted the strengths and limitations of daily gridded datasets, making us cautious about interpreting midnight weather readings.
3. **Reeves et al. (2009) “North American Regional Reanalysis (NARR) Data”**  
   Provided examples of working with netCDF climate data in large-scale research, guiding our netCDF-to-Parquet conversion.
4. **Running et al. (2004) “A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production”**  
   Demonstrated the importance of vegetation indices for fire modeling, inspiring our use of fuel load and cover type features.
5. **van Wagner (1987) “Development and Structure of the Canadian Forest Fire Weather Index”**  
   Deepened our understanding of the Fire Danger Index (bi), improving interpretation of its role in predictions.
6. **Mu et al. (2007) “Global Evapotranspiration Data from MODIS”**  
   Showed how evapotranspiration metrics (etr, pet) relate to fuel moisture, validating their inclusion as predictors.