ChatGPT

Key Predictor Categories for Short-Term and Long-Term Energy Consumption Forecasting

To accurately forecast electricity consumption for Foreca's needs, we must leverage a broad set of features that capture historical patterns, weather influences, calendar effects, and broader socio-economic trends. Below is a comprehensive list of the most relevant predictors, categorized by type, along with notes on whether they are more pertinent to short-term (48-hour) or long-term (12-month) forecasting.

Historical Load & Temporal Pattern Features

Past consumption data is one of the strongest predictors for future demand, as energy use often follows regular daily and weekly cycles. By incorporating lagged values and time-based patterns from the internal consumption series, we can capture these autocorrelations and seasonality:

- **Recent Lagged Consumption (Short-Term):** Including the last observed load and values from 24 or 48 hours ago (the same hour on the previous day and day before) helps the model recognize repeated daily patterns and any recent spikes or drops [1]. For example, yesterday's 6 PM consumption is a good predictor of today's 6 PM consumption, adjusting for other factors.
- **Weekly Seasonality (Short-Term):** The load one week ago (168 hours lag) or using Fourier terms to represent the *time of week* can capture recurring weekly usage cycles [2]. Many consumption patterns differ on weekdays vs weekends, and tend to repeat week over week.
- **Monthly/Yearly Lagged Values (Long-Term):** In monthly forecasting, using the consumption from the same month in previous years or recent months (e.g. consumption 12 months ago or last month) captures annual seasonality and trends. Electricity demand often has an annual cycle (higher in winter, lower in summer for heating-dominated regions) and a general growth or decline trend over years [3].
- **Rolling Averages/Trends (Both):** Features like the rolling mean of the past 7 days or past 30 days of consumption smooth out anomalies and highlight underlying trends. In long-term models, a moving average of recent monthly consumption can indicate momentum (e.g. a rising or falling usage trend beyond normal seasonality).

These historical and temporal features ensure the model accounts for intrinsic load cycles and recent behavior changes before considering external influences.

Calendar & Holiday Indicators

Calendar features help the model adjust for human routines and special events that drive energy usage. Electricity demand varies significantly by hour of day, day of week, and the presence of holidays or special events:

- **Hour-of-Day and Day-of-Week (Short-Term):** These capture diurnal and weekly patterns in consumption. For instance, hour-of-day indicates the daily cycle (with typical peaks in morning or early evening and lows at night), and day-of-week distinguishes weekdays (higher industrial/commercial load) from weekends [4]. Modeling time-of-week with cyclic encodings (e.g. sine/cosine or Fourier terms) is effective to represent these periodic patterns [5] [6].

- **Month or Season of Year (Long-Term):** In monthly forecasts, an indicator for month or season accounts for broad seasonal variations (e.g. winter vs summer). Finland's electricity use shows *strong seasonal variation peaking in winter* due to heating demand [3]. Including the month index or season helps the model know, for example, that January's baseline usage is typically higher than July's.
- **Public Holidays and Special Days (Short-Term):** Boolean flags for national holidays (and possibly days around them) are crucial, as holidays can drastically reduce industrial and commercial electricity consumption and shift residential usage patterns [6]. For example, Christmas Day or Midsummer in Finland will have atypically low demand during what would otherwise be a weekday. The model should know those timestamps are holidays to avoid over-predicting consumption. (In practice, one can use an open-source calendar of Finnish public holidays to create this feature.)
- **Workday/Weekend Count (Long-Term):** For monthly aggregation, the number of working days vs. weekends/holidays in a month can serve as a feature. Months with fewer workdays (or extended holiday periods like July summer vacations or December holidays) tend to have lower total consumption. An open data source (e.g., Finnish calendar) can provide how many weekdays vs weekends a given month has.
- **Special Events or Seasonal Hours:** If relevant, one might also include events like daylight savings time changes or major national events that impact electricity use. In Finland, the transition in daylight hours is extreme across seasons; while not a "holiday," it influences behavior. For example, very short daylight in winter can increase lighting demand, so implicitly capturing this through date/season or a "daylight hours" feature can be useful (more on daylight under weather).

Calendar features are mostly deterministic and open data (holidays schedules, etc.), and they ensure the model differentiates ordinary days from special ones that alter consumption patterns.

Weather and Environmental Factors

Weather is one of the most critical external drivers for electricity consumption, especially in a climate like Finland's where heating demand is high. Integrating meteorological data (which is available from open sources like the Finnish Meteorological Institute) can greatly improve forecast accuracy [4]:

- **Temperature (Short-Term & Long-Term):** Ambient temperature has a dominant effect on energy use for heating and cooling. In Finland, **electricity consumption correlates extremely well with the average temperature** [7]. For short-term hourly forecasting, using forecasted temperatures for the next 48 hours at the relevant locations is ideal (e.g. hourly temperature predictions from a weather API). As temperature drops in winter, electric heating loads soar; conversely, mild weather reduces demand. For long-term (monthly) forecasts, one can use aggregated temperature metrics – such as **heating degree days (HDD)** and **cooling degree days (CDD)** for each month – to quantify how far temperatures deviate from comfort levels. For example, total HDD in a month (based on open climate data) can serve as a feature indicating heating demand that month. If future weather is unknown, using 30-year climate normals or average HDD/CDD for those months is a reasonable assumption.
- **Humidity & Heat Index (Short-Term):** Humidity combined with temperature (heat index) affects cooling demand in summer. While Finland's cooling load is smaller than heating, high humidity in warm weather can drive air conditioning usage. Including humidity or dew point from open data can

refine short-term models [8]. In winter, humidity has less direct effect on heating, but extreme low humidity might correlate with very cold air masses.

- **Precipitation and Snow (Short-Term):** Rain or snow can have indirect effects on consumption. For instance, overcast rainy days are darker, potentially increasing lighting use during daytime, and heavy snowfall or storms might disrupt normal routines (people staying indoors, slight changes in consumption). These effects are secondary, but precipitation data (rainfall, snowfall) from weather sources could marginally improve accuracy in short-term models [8]. For long-term monthly predictions, one might not include precipitation explicitly, as its effect on total monthly energy is less pronounced compared to temperature.
- **Wind Speed (Short-Term):** Windy conditions can influence heating demand due to wind chill (making cold temperatures feel colder and buildings lose heat faster). If available, wind speed and gust data can be used as features to adjust heating-related consumption in short-term forecasts [8]. Additionally, severe storms (high winds) might cause preventive shutdowns or anomalies in usage, but those are hard to predict.
- **Daylight Hours / Solar Radiation (Both):** The amount of daylight influences lighting and, to a degree, heating. Finland's high latitude means winter days are very short and summers have very long days. Including a feature for **day length** or average solar radiation can help capture this effect. In winter, fewer daylight hours mean more lighting demand (and slightly more heating since the sun's radiant heat is minimal), whereas in summer, long days reduce lighting needs. For short-term forecasts, day length is implicit in the date/time (but a model might benefit from a feature like "sunset time" or an indicator for daylight vs dark hours). For monthly forecasts, one could use average daylight hours in that month as an open data feature to explain differences in lighting usage between months.
- **Local Weather by Region/Group:** Given each user group has a region/municipality, matching weather data to each group's location will yield more accurate features. For example, using the temperature in Helsinki for a southern coastal customer group vs. temperature in Lapland for a northern group captures regional demand drivers. The model could include multiple locations' weather if groups are spread out [9]. Finnish Meteorological Institute provides open observations and forecasts for various locations which can be leveraged.

Incorporating weather features ensures the forecast responds to imminent changes (a coming cold front or heatwave) in short-term and captures seasonal climate impacts in long-term. These data are openly available and highly relevant for energy use prediction [4].

Electricity Market & Price Signals

Electricity pricing and market conditions can influence consumption, especially in scenarios of demand response or price-sensitive usage. The provided day-ahead price data (Nord Pool spot prices, which are publicly available) is a valuable predictor, but it must be used carefully given forecasting horizons:

- **Day-Ahead Electricity Price (Short-Term):** The day-ahead market price for power is known for the next 24 hours and can be included as a feature for the first day of the 48-hour forecast. Price affects consumer behavior **after the market results are published**, as some large consumers may adjust consumption to capitalize on cheaper hours or reduce load during expensive hours [10]. For instance, industries or smart-grid enabled customers might shift discretionary usage from peak price hours to off-peak if prices are high. The model can learn that extremely high prices often coincide with slightly lower demand than usual (and vice versa) due to this elasticity.

- **Price for Second Day (Longer Horizon):** Since day-ahead prices for 25–48 hours ahead are not known in reality, one approach is to exclude price for the second day or use an approximation (such as assuming it equals the first day's price pattern or using a forecasted price if available). It's important to note explicitly that for hours 25-48 the model cannot rely on actual prices, to simulate a real scenario.
- **Market Bidding/Consumption Volumes (Short-Term):** Aside from price, the day-ahead market publishes information on electricity purchase bids and volumes. **Market purchase volume** can correlate with consumption since it reflects how much energy was scheduled for delivery [10]. In advanced usage, features like the *difference between forecasted consumption and actual consumption of yesterday* or *grid imbalance* could hint if today's usage might be higher or lower. The Aalto research introduced multiple market-based features (both price and volume from bid curves) to improve Finland's load forecasts [11]. Such data is open (e.g., Nord Pool or Fingrid publishes market results) and can be included if relevant to Foreca's context.
- **Price Change Indicators:** It has been observed that the *change* in price or other market values from the previous day can be more predictive than the absolute price [12]. For example, a sudden jump in price compared to yesterday might signal an unusual supply/demand situation that also reflects in consumption changes. Thus, features like *price delta from previous day* (for known hours) or a flag for "price spike" can be considered.
- **Tariff or Demand Response Programs:** If any user groups are on time-of-use tariffs or demand response agreements, knowing those tariff schedules (often public information) can help. For instance, if residential customers have cheaper night rates, the model could anticipate higher usage at night. In Finland, many households have two-rate meters (day vs night tariff) [13], which partly ties into price signals. While our dataset segmentation might cover this via product type metadata, explicitly modeling a feature for "cheap rate period vs expensive period" could improve short-term load shape predictions.

Market and price data are generally more immediately useful for short-term forecasts. They add a layer of behavioral insight (consumers reacting to market conditions) on top of purely weather- or calendar-driven demand. Indeed, recent forecasting models for Finland have started using such **market features (like day-ahead price and bought energy volume)** to enhance accuracy [14]. For long-term (monthly) forecasting, price is harder to include because future prices 12 months out are unknown and demand elasticity over that horizon is usually low – however, one could incorporate an assumption or scenario of average price levels if needed, or simply rely on the notion that price effects average out in seasonal totals.

Demographic and Economic Factors

For longer-term consumption trends (and to some extent short-term anomalies), socio-economic drivers play a role. These features capture changes in population, economic activity, and other broad trends that influence energy usage:

- **Population and Customer Growth (Long-Term):** An increase or decrease in the number of consumers will change total consumption. If open data on population by region or number of households is available (e.g., from Statistics Finland), it can be used to adjust forecasts, especially over a 12-month horizon. For example, a growing population or new housing development in a region would raise residential electricity demand. In our case, if the 112 user groups correspond to distinct customer sets, the relative size of each group (number of customers) could be a feature. Population tends to change slowly, so it's more relevant in long-term trend forecasting [15].

- **Economic Activity / Industrial Production (Long-Term):** Economic indicators like GDP growth, industrial production index, or business output can influence electricity consumption, particularly for commercial and industrial segments [15]. For instance, if the economy is booming, factories run at higher capacity and use more power; in a recession, industrial electricity use may drop. Open-source data for Finland's monthly industrial output or economic indices can enrich the model. These are especially useful for the 12-month forecasts, as they account for demand changes not explained by weather or season alone (e.g. a new industrial plant coming online or an economic slowdown).
- **Energy Efficiency Trends (Long-Term):** Over time, appliances and processes become more energy-efficient, potentially reducing consumption growth. While hard to quantify directly, proxy variables like the penetration rate of energy-efficient technologies or **policy changes** (e.g., new efficiency standards, subsidies for insulation, etc.) could be considered. One might incorporate a gentle downward trend or a dummy for years after a major policy. If there are known upcoming changes (all open data), e.g., a scheduled closure of a large factory or an opening of one, those could be factored in as scenario adjustments to the forecast.
- **Electric Vehicle (EV) and Electrification Impact (Long-Term):** As Finland's transportation electrifies, charging EVs will increase electricity demand. If data on EV adoption (e.g., number of EVs registered monthly) is available openly, it could be included to project additional load, particularly in residential segments. Similarly, trends like heat pumps adoption for heating (replacing oil/wood heating with electric heat pumps) would raise electrical consumption in winters. Such data may be obtainable from energy authorities or statistical agencies and would refine long-term forecasts of monthly consumption in sectors undergoing electrification.
- **Geomagnetic or Other Environmental Indices:** (This is a niche factor, but worth noting given Finnish research interest.) A recent study even found that **geomagnetic activity (energetic particle precipitation)** has a subtle influence on Finland's winter temperatures and thus electricity demand [7]. While not a typical feature, it underscores that any reliable climate or environmental indicator that helps predict temperature can indirectly serve the load forecast. However, for practical purposes, including such variables would be an advanced step and only if the effect is significant and data openly available (e.g., indices from NOAA).

Demographic and economic features are more crucial for long-term forecasts, where **"broader economic, demographic, or environmental trends"** come into play [16]. In short-term (48h), these factors are mostly constant and thus negligible. But for predicting monthly consumption a year out, accounting for growth or decline in demand drivers beyond weather (e.g. population changes, economic growth, new infrastructure) can improve accuracy [15].

Customer Segment & Regional Attributes

The provided metadata about each user group (such as region, municipality, customer segment, and product type) can be leveraged as features or to tailor the model per group. These attributes help the model understand structural differences in consumption patterns:

- **Regional Location:** Knowing the macro-region or municipality of a user group allows pairing with relevant weather data and capturing regional usage patterns. For example, a group in Northern Finland will have a different weather profile (colder winters) and possibly a different consumption baseline than one in Southern Finland. Region can be included as a categorical feature (one-hot encoded or similar) in a combined model, so the model can learn regional offsets and sensitivities. It also enables use of **regional weather features** (temperature, etc., as discussed) specific to that location.

- **Customer Segment (Residential vs Commercial vs Industrial):** The segment or type of customer influences load shape. Industrial users tend to have high weekday usage and possibly lower on weekends or during holiday shutdowns, while residential users peak in mornings/evenings and might use more on weekends when people are home. By using the segment metadata, one can either build separate models for each segment or include it as a feature in a unified model to adjust behavior. For instance, a binary feature for "industrial customer" might help the model learn a stronger weekday/weekend difference for that group. Segment information is internal, but it's given data that greatly enhances feature representation.
- **Product Type/Tariff:** If the product type indicates the tariff structure (e.g., fixed rate vs time-of-use tariff), this can inform how price-sensitive that group is. A group on a real-time pricing contract might have load that responds to price (so the price feature will matter more), whereas a fixed-rate group's consumption might be more weather-driven. Including product type as a feature or stratifying models by it can capture these differences.
- **Static Attributes:** Other static attributes like connected load capacity, building types (if known), or urban vs rural could also be considered. For example, urban areas might have more electric public transport or different usage patterns than rural. While not explicitly given, some might be inferred from the region or segment data (municipality size could hint at urban/rural). Such features help in long-term modeling to differentiate growth rates or base usage levels.

By categorizing and utilizing these internal metadata features, we ensure that the model's other predictors (weather, calendar, etc.) are interpreted in the proper context for each user group. For example, 0°C and snow in Lapland (where it's normal and homes are well-insulated) might have a different impact on consumption than 0°C in coastal Southern Finland. Including segment and regional features allows the model to learn these nuances.

In summary, the best predictor set for the energy consumption forecasting task spans multiple categories: **historical load patterns, calendar/seasonal effects, weather variables, market price signals, and macro-level trends**, as well as leveraging the provided metadata to customize predictions for different groups. Short-term models (48h ahead) will rely more on high-frequency factors like recent load, hourly weather forecasts, and known next-day prices [4] [10]. Long-term models (12 months ahead) emphasize lower-frequency drivers like seasonal averages, month-of-year, and economic/demographic trends to capture year-ahead changes [16] [15]. By combining these features – all obtainable from the given data and open sources – we can cover the key predictors that drive electricity consumption in Finland, leading to robust forecasts for both the next days and the next year.

Sources:

- IBM Think Blog – *What is load forecasting?* (Factors influencing electricity use: weather, time, calendar, demographics) [4]
- Vehtari (2025) – *Forecasting day-ahead electricity consumption in Finland* (Feature categories used: historical load, market data, weather, calendar/holiday, time-of-week) [14] [17]
- Aalto Thesis (2025) – Discussion on using market price and holiday indicators as features in load forecasting [6] [10]

- Scientific Reports (2023) – Correlation of Finland’s electricity consumption with temperature and seasonal patterns (peak in winter, trend with industry changes) [7] [3]
- Li & Chen (2019) – *Monthly Electricity Consumption Forecast* (Long-term factors: *“population density, economic growth, power facilities, and climate factors”* influence power demand) [15]
- Milvus AI Reference – *Short-term vs Long-term forecasting* (Short-term focuses on recent cyclical patterns; long-term incorporates broader economic/demographic trends) [16]

---

[1] [2] [5] [6] [8] [9] [10] [11] [12] [14] [17] Forecasting day-ahead electricity consumption in Finland
https://aaltodoc.aalto.fi/bitstreams/01572f55-13c9-404c-9da2-48b06711ab43/download

[3] [7] Electricity consumption in Finland influenced by climate effects of energetic particle precipitation | Scientific Reports
https://www.nature.com/articles/s41598-023-47605-8?error=cookies_not_supported&code=240f9180-f808-4f26-8c75-b031ad46d7f1

[4] What Is Load Forecasting? | IBM
https://www.ibm.com/think/topics/load-forecasting

[13] Basic Electricity (2-rate) | Helen
https://www.helen.fi/en/electricity/electricity-products-and-prices/basic-electricity-2-rate

[15] Monthly Electricity Consumption Forecast Based on Multi-Target Regression
https://www.scirp.org/journal/paperinformation?paperid=93833

[16] What is the difference between short-term and long-term forecasting?
https://milvus.io/ai-quick-reference/what-is-the-difference-between-shortterm-and-longterm-forecasting