

# JOB SILL DEMAND

## TREND ANALYSIS & SHORT - HORIZON FORECASTING

---

Name: Kirti Gupta  
Course: Minor in AI  
Date: August 2025

### ABSTRACT

This project analyzes job skill demand trends using LinkedIn job postings. The pipeline includes data cleaning, salary harmonization, time-series construction, baseline forecasting (Naive, MA, SES, ARIMA), Prophet modeling, and a compact classification task predicting short-term skill demand shifts. Results are exported as reproducible artifacts, including metrics, forecasts, and polished plots.

### DATA & PRE - PROCESSING

**Source:** postings.csv, jobs/job\_skills.csv, mappings/skills.csv from **Kaggle:** linkedin-job-postings.

Cleaning included deduplication by job\_id, filling missing values, clipping outliers, salary harmonization (annual USD conversion), and log-transform for salary EDA. Skills were joined in long format. Time series were constructed as counts of unique job IDs/day. Features engineered included rolling means, differences, z-scores, and MA ratios.

### METHODS

**Baselines:** Naive, MA(k), SES, ARIMA(1,1,1) with MAE/RMSE/sMAPE/MASE metrics.

**Prophet:** Daily models with weekly seasonality, additive mode, and changepoint tuning. Weekly Prophet attempted where possible.

**Classification:** Logistic Regression pipeline with accuracy-oriented threshold tuning and fallback walk-forward CV when hold-out was tiny.



## RESULTS

**Baselines vs Prophet:** SES often outperforms Prophet on small, noisy series (e.g., Engineering: SES MAE $\approx$ 1104 vs Prophet $\approx$ 1145).

**Adaptive CV:** SES/MA(7) robust across multiple skills; ARIMA/Naive weaker.

**Prophet daily:** Writing/Editing MAE $\approx$ 123, Quality Assurance $\approx$ 150, Accounting $\approx$ 471, Engineering $\approx$ 1145, IT $\approx$ 2298.

**Classification:** Engineering achieved  $\approx$ 0.67 accuracy (walk-forward CV) with tuned thresholds.

## DISCUSSION

SES and MA(7) provide reliable baselines. Prophet is competitive on steadier skills but less effective on volatile ones. Classification is limited by data sparsity but demonstrates threshold tuning.

**Limitations:** Sparse series, weekly nonzero scarcity, noisy classification labels.

**Next Steps:** Aggregate to weekly series, enrich features (holidays, regions), pooled/hierarchical models, and calibrated classification.

## REPRODUCIBILITY & ASSETS

**Artifacts:** metrics\_multi\_skill\_D.csv, advanced\_adaptive\_results.csv, prophet\_daily/weekly metrics & forecasts, daily\_skill\_features.csv, engineering-focused comparison CSVs and plots, and PROJECT\_CARD.json.

**Notebook:** End-to-end pipeline saving outputs to artifacts folder.

## CONCLUSION

Short-horizon skill demand forecasting shows SES/MA(7) as robust benchmarks. Prophet is useful with richer signals.

Classification accuracy is constrained but improved with threshold tuning.

All results are reproducible with saved artifacts and documented pipeline.

