OLABAM · DATA SCIENCE

Digital Burnout & Productivity Analytics — Project Report

Author: Olamide Bamigbola · **Type:** End-to-end regression & deployment · **Model:** XGBoost Regressor

1. Executive summary

This project investigates how digital behaviour and wellbeing influence workplace productivity, and builds a deployable model that predicts a continuous **productivity score (0–100)** from a person's daily habits and state.

Working from a **5,000,000-row** behavioural dataset, the pipeline cleans the data, engineers domain-informed features, benchmarks six regression algorithms, and selects an **XGBoost** model that explains **86.6%** of the variance in productivity (**R^2 = 0.866, RMSE ≈ 7.86**). The model is served through a Flask API and an interactive web interface featuring a colour-coded speedometer gauge.

The central finding is consistent across every stage of analysis: **deep work and the capacity to focus are the dominant, robust drivers of productivity**, while burnout, stress and emotional exhaustion are the main drags. Surface-level "screen time" metrics, by contrast, have only weak and largely non-linear relationships with output.

2. Problem statement

Organisations increasingly worry about "digital burnout" — the productivity cost of fragmented attention, doomscrolling and always-on devices. The goal of this project is twofold:

1. **Quantify** which behavioural and wellbeing factors most strongly relate to productivity.
2. **Predict** an individual's productivity score from those factors, in a form that can be served live.

This is framed as a **supervised regression** problem with `productivity_score` as the target.

3. Dataset

- **Size:** 5,000,000 rows × 34 columns (≈1.3 GB in memory).
- **Composition:** 29 numerical features, 5 categorical features.
- **Targets available:** `productivity_score` (continuous, 0–100), `productivity_category` (Low / Medium / High), and `burnout_risk`.

Features span five conceptual groups:

Group	Example features
Digital behaviour & screen usage	`daily_screen_time`, `social_media_hours`, `doomscrolling_duration`, `app_switch_frequency`, `notification_count`, `smartphone_unlocks`
Focus & deep-work habits	`focus_sessions`, `deep_work_hours`, `distraction_frequency`, `concentration_score`, `task_completion_rate`
Health & wellbeing	`sleep_hours`, `sleep_quality`, `caffeine_intake`, `physical_activity`, `stress_level`
Work environment	`workspace_quality`, `meeting_hours`, `internet_stability`, `remote_work_days`
Psychological state	`motivation_level`, `mental_fatigue`, `emotional_exhaustion`, `work_satisfaction`, `mental_state`

4. Data cleaning

- **Missing values:** below 3% overall, concentrated in `social_media_hours`, `deep_work_hours`, `sleep_hours` and `motivation_level`. Given the low proportion, **mean imputation** was applied.
- **Duplicates:** none found.
- **Memory optimisation:** integer and float columns were downcast to smaller dtypes, reducing memory from ~1.3 GB to ~1.17 GB and making the 5M-row workflow tractable.

5. Exploratory data analysis

Target distribution. `productivity_score` is mildly left-skewed (skew ≈ −0.45), concentrated in the mid-to-high range. The categorical label is imbalanced:

Category	Count	Mean score
High	2,374,159	90.09
Medium	1,784,357	62.85
Low	841,484	37.05

Correlation findings (vs. `productivity_score`):

- **Strongest positive:** `deep_work_hours` (**0.555**), `task_completion_rate` (0.356), `concentration_score` (0.347), `focus_sessions` (0.230), `motivation_level` (0.229).
- **Strongest negative:** `burnout_risk` (**−0.378**), `emotional_exhaustion` (−0.182), `social_media_hours` (−0.140), `stress_level` (−0.135).
- **Weak / negligible:** raw screen-usage counts (notifications, unlocks, daily screen time) showed near-zero linear correlation, implying their influence is non-linear or minimal.

Group-level insights:

- **Focus & deep work** is the most predictive cluster; deep work alone correlates 0.555 with productivity, and that relationship holds across all work modes (Office / Hybrid / Remote) and all sleep-quality levels — making it a stable, robust signal.
- **Health & wellbeing** factors show weak *direct* effects, with emotional exhaustion and stress the only meaningful negative drivers.
- **Work environment** matters mainly through motivation and workspace quality; meetings, internet stability and remote-work days were negligible.
- **Mental state** emerged as the most significant categorical differentiator of productivity.

6. Feature engineering

Thirty-two composite features were derived to capture interactions that raw columns miss, expanding the set from 35 to 67 features. Notable examples:

- **Digital strain:** `total_distraction_exposure`, `phone_checking_compulsion`, `social_media_engagement`, `notification_intensity`, `digital_wellness`.
- **Biological readiness:** `sleep_deficit`, `sleep_effectiveness`, `caffeine_dependence`, `daily_energy_balance`, `stress_resilience`.
- **Psychological readiness:** `emotional_stability`, `burnout_trajectory`, `motivation_durability`, `psychological_readiness`.

- **Work environment & interaction:** `deep_work_efficiency`, `work_environment_quality`, `productive_time_available`, `cognitive_overload`, `focus_support_score`, `productivity_capacity`.

Several engineered features ranked among the most predictive: `deep_work_efficiency` (r = 0.433), `psychological_readiness` (0.266), `remote_work_sustainability` (0.258) and `emotional_stability` (0.236) — validating the domain-driven approach.

7. Preprocessing & feature selection

- **Encoding:** one-hot encoding for `occupation`, `work_mode` and `device_usage_type` (drop-first); ordinal mapping for `mental_state` (Burnout=1 → Focused=4) and the category labels (Low=0, Medium=1, High=2).
- **Feature selection:** features with |correlation| < 0.03 to the target were dropped (30 weak features removed); pairs with |correlation| > 0.95 were checked for multicollinearity. The final set contained **39 predictors**.
- **Split:** 70% train / 15% validation / 15% test → 3,502,000 / 748,000 / 750,000 rows. Target means matched across splits (≈71.4), confirming a clean, representative partition.
- **Scaling & imputation:** `StandardScaler` standardised the features; a `SimpleImputer` (mean) handled residual missing values introduced during transformation.

8. Modelling

Six regression algorithms were benchmarked on a **1,000,000-row stratified sample** (stratified on binned target) for efficient comparison, then evaluated on the full held-out test set.

Model	Test R^2	Test RMSE	Overfit gap (train−test R^2)	Train time
XGBoost	**0.866**	**7.865**	0.0229	39.5s
Gradient Boosting	0.857	8.118	0.0083	304.7s
Linear Regression	0.841	8.584	0.0013	1.8s
Random Forest	0.815	9.254	0.0779	187.7s
Decision Tree	0.711	11.563	0.0528	8.3s
AdaBoost	0.702	11.743	0.0019	58.7s

Selected model — XGBoost (`n_estimators=300, max_depth=6, learning_rate=0.05`). It delivered the best accuracy with a small overfit gap and trained in under a minute — an excellent accuracy/efficiency trade-off. The strong performance of plain Linear Regression also confirms a large, genuinely linear signal in the engineered features.

9. Model interpretability (SHAP)

SHAP analysis of the XGBoost model ranked the most influential features as:

1. `deep_work_hours`
2. `task_completion_rate`
3. `concentration_score`
4. `burnout_risk`
5. `motivation_level`
6. `focus_sessions`
7. `workspace_quality`
8. `social_media_hours`
9. `app_switch_frequency`

This is consistent with the EDA: higher deep work, task completion, concentration and motivation push predictions up, while higher burnout, social-media hours and app-switching push them down. The model's logic is interpretable and aligns with domain intuition.

10. Deployment

- The trained model is serialised to `XGBoost_model.pkl` with **joblib**.
- A **Flask** API (`app.py`) exposes `POST /predict`, returning a JSON `prediction`. **flask-cors** allows the browser front-end to call it.
- The service is deployed on **Render**; the interactive front-end (an OLABAM-branded predictor with a colour-coded speedometer gauge and light/dark theme) is served as a static site via **GitHub Pages** at **olabam.com**.
- The public prediction endpoint accepts nine of the most influential behavioural inputs (`burnout_risk`, `social_media_hours`, `app_switch_frequency`, `distraction_burden_work`, `deep_work_hours`, `focus_sessions`, `concentration_score`, `task_completion_rate`, `motivation_level`).

11. Limitations & future work

- **Reduced-input endpoint.** The live API fills only the nine headline features and defaults the remaining model inputs to zero. Exposing the full feature set — or supplying realistic population defaults — would improve prediction fidelity for end users.
- **Synthetic-style data.** The dataset's near-zero correlations among raw inputs suggest a partly synthetic generation process; relationships should be validated on real-world observational data before operational use.
- **Class imbalance.** The High-productivity class dominates; if the categorical task is pursued, resampling or class weighting is advisable.
- **Cold starts.** On Render's free tier the API sleeps when idle, so the first request can take up to a minute — a paid tier or a keep-alive ping would remove this.
- **Future directions:** hyperparameter tuning (Optuna), confidence intervals on predictions, a companion burnout-risk classifier, and per-prediction SHAP explanations surfaced in the UI.

12. Conclusion

The project delivers an accurate, interpretable and deployed model for predicting productivity from behavioural and wellbeing signals. Its clearest, most actionable insight is that **protecting deep, focused work time and managing burnout matter far more than simply cutting screen time** — a conclusion that survived correlation analysis, group-wise breakdowns, and SHAP attribution alike. With **R^2 = 0.866** on unseen data and a working live interface, the system demonstrates a complete path from raw data to a usable, explainable product.