

Intent Prediction in Online Dating:

Algorithmic Alignment & Causal Drivers of Family Planning

 python  pandas  spaCy

  Shap

Executive Summary

Optimizing Match Relevance by Decoupling Interest from Intent



Demographics fail to capture the nuance of user intent.

Demographics provide **79% baseline** accuracy for family planning but miss the **child-free** vs. **not ready** distinction.



Romance signals are not a proxy for family planning.

Users seeking “love” and “connection” showed **no meaningful increase** in family planning, highlighting an inverse of algorithmic assumptions.



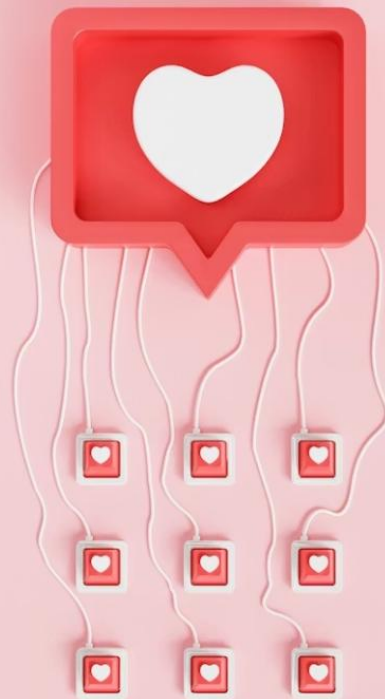
Psychographic personas provide improved matching.

Socialite behaviors **increase** family planning by **+18.5%**, while **Techie** interests **decrease** family planning by **-19.8%**.



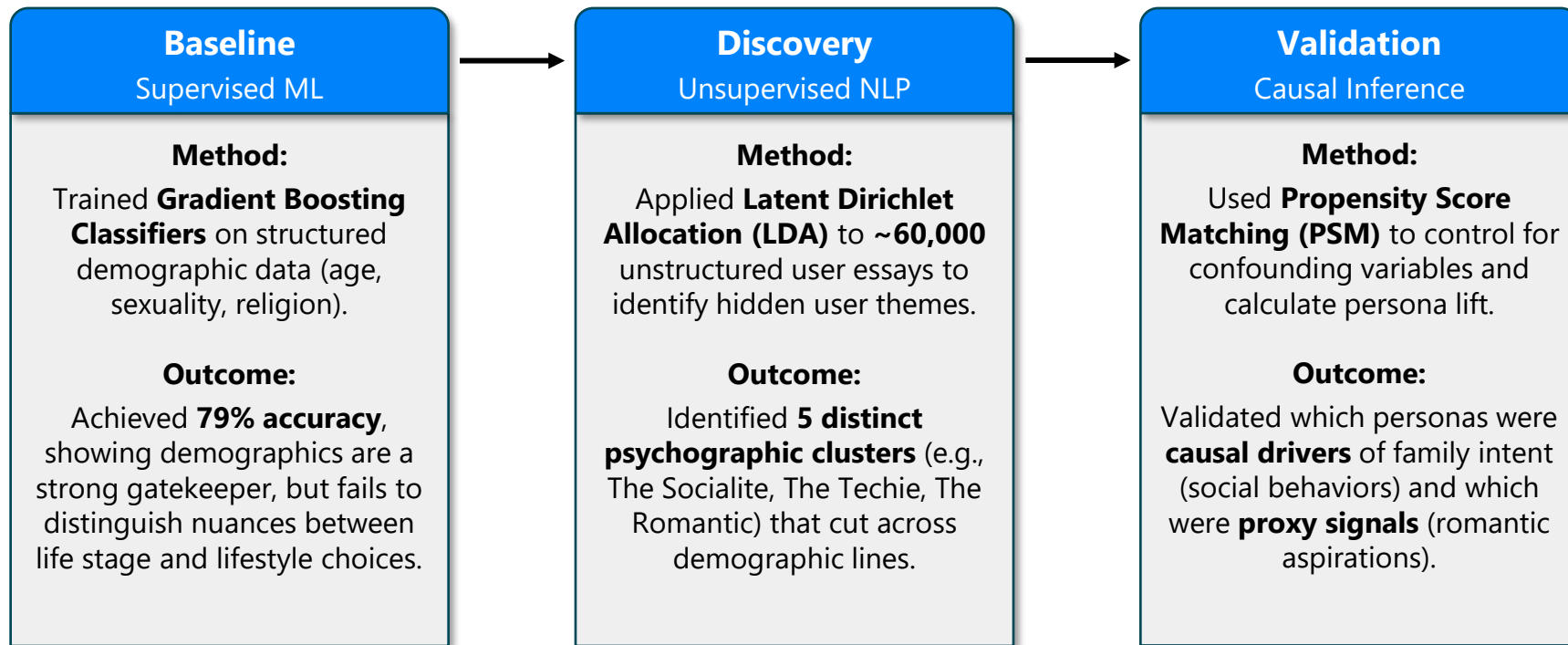
Algorithmic alignment requires distinct tracks.

Decouple **interest-based** matching from **intent-based** matching to reduce false positives and avoid nudging users toward incompatible partners.



From Data to Insight: The Research Pipeline

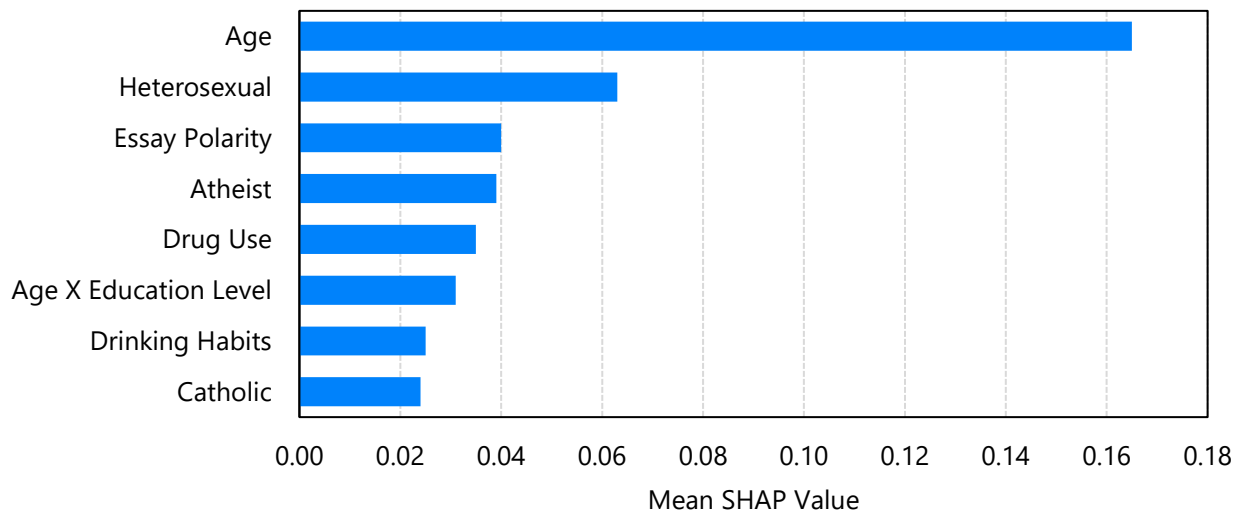
Triangulating User Intent via Machine Learning, Natural Language Processing, and Causal Inference



Demographics Miss Lifestyle Choice Nuance

Baseline Gradient Boosting Model: SHAP Feature Importance Analysis

Global Feature Importance



Demographic Predictors

(> .02 SHAP Value)

Life Stage:

Age & Age X Education Level

Identity:

Sexual Orientation & Religion

Emotional Disposition:

Essay Polarity

Habits:

Drug Use & Drinking Habits

The model focuses too heavily on **Life Stage** and misses **Lifestyle Choices**

A 35-year-old **child-free** user is **identical** to a 35-year-old **family-seeker**, leading to high-friction matches.

Unlocking Psychographics: Language is a Behavior

Latent Dirichlet Allocation (LDA) Topic Modeling of User Essays



The Artist

Keywords:
Art, Film, Band

Signal: Non-traditional lifestyle; values expression, creativity, and aesthetics



The Techie

Keywords:
Game, Video, Computer

Signal: Niche interests; highly specific media and tech culture compatibility



The Socialite

Keywords:
TV, Sports, Dinner

Signal: High social activity; values community and shared experiences



The Professional

Keywords:
Work, Home, Business

Signal: High ambition; prioritizes stability and career trajectory



The Romantic

Keywords:
Relationship, Soul, Heart

Signal: Intimacy-focused; explicitly seeking deep emotional connection

Lifestyle Choices Drive Family Planning

Validating Causal Drivers via Propensity Score Matching (PSM)

Propensity Score Matching

Challenge:

Are certain personas (e.g., Socialites) more family-oriented, or just older?

Method:

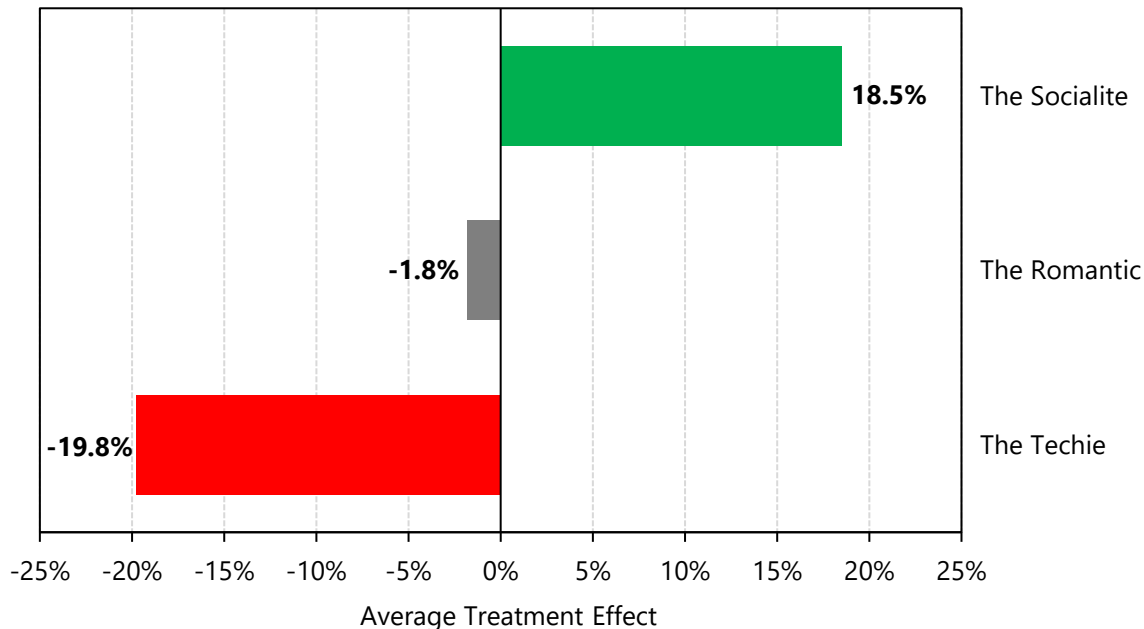
Create statistical “twins” for every user (a demographically identical “non-persona”) to isolate causal effects.

Insight:

Lifestyle choice is a significant causal driver.

- **Socialites:** Community focus aligns with family goals.
- **Techies:** Niche interests compete with family planning.
- **Romantics:** “Love” signals are noise, proving intent is distinct from connection.

Causal Lift on Family Planning



From Insight to Impact: Strategic Recommendations

Optimizing Product Features Based on Psychographic Drivers



The Algorithm

Decouple Interest from Intent

Insight:

Niche interests **actively suppress** family intent signals.

Recommendation:

Downweight **interest-based** matching. Prioritize **latent psychographics** as a primary feature.



The Onboarding

"Lifestyle" Signal Onboarding

Insight:

Demographics achieve **79% accuracy** but miss nuance.

Recommendation:

Replace static demographic filters with a **visual lifestyle quiz**. Let users self-select into specific psychographic tracks (e.g., Socialite, Romantic).



The Growth

Segmented Value Proposition

Insight:

"Romantics" seek **connection**, not commitment.

Recommendation:

Stop retargeting "Romantic" users with ads about family. Shift focus to **"deep connection" messaging** to align with actual intent.

Execution Strategy: Validation & Rollout

Translating Insights into Measurable Product Impact

A/B Test Design

Hypothesis:

Users who complete a **lifestyle quiz** during onboarding will show higher day-30 retention than standard onboarding users.

Test Setup:

Treatment: Lifestyle Quiz

Control: Demographic filters

Traffic: 5% of new user traffic

Success Metrics

Primary:

Day-30 Retention.
(Do better matches keep users?)

Secondary:

Conversation Initiation Rate
(Do matched users engage?)

Guardrail:

Onboarding Completion Rate
Match Volume

Phased Rollout

Phase 1:

Backtest on historical data ✓

Phase 2:

Live A/B Test (5% traffic)

Phase 3:

Ship if day-30 retention shows
≥ **2% lift**