

عنوان: پیش‌بینی‌کننده‌های ترک سیگار در ایالات متحده: یک تحلیل یادگیری ماشینی که پارادوکس اجتماعی-اقتصادی را در داده‌های (2018-2023) BRFSS آشکار می‌کند.

Title: Predictors of Smoking Cessation in the U.S.: A Machine Learning Analysis

Revealing the Socioeconomic Paradox in BRFSS Data (2018-2023)

نویسندگان و گردآوردگان: آرمین خوجوی، حامد حسامی و مهدیه ابراهیمی

Keywords: Smoking Cessation, Machine Learning, BRFSS (Behavioral Risk Factor Surveillance System), Population Health

چکیده (Abstract)

زمینه و هدف: ترک موفق سیگار، به عنوان یکی از چالش‌های اصلی بهداشت عمومی، تحت تأثیر مجموعه‌ای پیچیده از عوامل فردی و محیطی قرار دارد. این تحقیق با هدف ارائه درکی عمیق، چندوجهی و پایدار از این عوامل، به تحلیل و مدل‌سازی پیش‌بینی‌کننده‌های ترک سیگار در بزرگسالان با استفاده از داده‌های وسیع پیمایش BRFSS در یک دوره شش‌ساله (۲۰۱۸-۲۰۲۳) می‌پردازد تا شواهد داده‌محور برای طراحی مداخلات شخصی‌سازی شده فراهم آورد.

روش‌ها: این مطالعه طولی، با به‌کارگیری مجموعه‌ای از مدل‌های یادگیری ماشین (از جمله رگرسیون لجستیک، LightGBM، CatBoost، و XGBoost) برای پیش‌بینی ترک سیگار بهره گرفت. عملکرد این مدل‌ها با استراتژی "تمام ویژگی‌ها" در مقایسه با روش "کاهش ویژگی بر اساس هم‌خطی (VIF)" ارزیابی شد. برای افزایش دقت و پایداری پیش‌بینی، رویکردهای مدل‌سازی جامع شامل **تجمیع داده (Data Pooling)**، **آنسامبل زمانی (Temporal Ensemble)** و به ویژه **روش انباشت (Stacking Ensemble)** نیز پیاده‌سازی شدند. تأثیر و جهت‌گیری هر عامل با استفاده از تحلیل SHAP ارزیابی و روند پایداری آن در طول زمان بررسی گردید.

یافته‌ها: مدل **Stacking Ensemble**، با کسب بالاترین عملکرد کلی (F1-Score: 0.7446، AUC: 0.7881)، برتری رویکردهای جامع را در پیش‌بینی ترک سیگار نشان داد. مهم‌ترین یافته، پایداری شگفت‌انگیز الگوهای

کلیدی پیش‌بینی‌کننده در طول دوره مطالعه بود. تحلیل SHAP به طور مداوم و پایدار، سن بالاتر را به عنوان قوی‌ترین عامل افزایش‌دهنده احتمال ترک موفق سیگار شناسایی کرد. همچنین، متأهل بودن، سطوح بالاتر درآمد و تحصیلات، BMI بالاتر و وزن بیشتر به طور مداوم با احتمال بیشتر ترک سیگار مرتبط بودند. از سوی دیگر، تشخیص بیماری‌های جدی جسمی (مانند COPD، سکتة مغزی) و اختلال افسردگی تشخیص داده شده به عنوان "نقطه بحرانی" عمل کرده و شانس ترک سیگار را افزایش دادند. با این حال، وضعیت سلامت عمومی ضعیف‌تر و مشکلات سلامت روان نامناسب (به‌ویژه از سال ۲۰۲۰/۲۰۱۹ به بعد) به طور پیوسته به عنوان مانع مهمی برای ترک موفق سیگار شناسایی شدند. انجام منظم فعالیت فیزیکی و چکاپ‌های پزشکی نیز

با احتمال بالاتر ترک سیگار همراه بود. در مدل‌های پایه، رویکرد استفاده از "تمام ویژگی‌ها" به طور مداوم عملکرد بهتری را نسبت به ویژگی‌های منتخب VIF ارائه داد، هرچند که در تمامی مدل‌ها یک **سوگیری پایدار به سمت پیش‌بینی کلاس اکثریت** ("ترک موفق") مشاهده شد.

نتیجه‌گیری: این تحقیق نشان می‌دهد که استفاده از مدل‌های یادگیری ماشین غیرخطی و رویکردهای ترکیبی، به ویژه Stacking Ensemble، نتایج برتری را در پیش‌بینی ترک سیگار و شناسایی عوامل پایدار ارائه می‌دهد. با توجه به ثبات چشمگیر پیش‌بینی‌کننده‌های اصلی در طول دوره مطالعه (حتی در مواجهه با رویدادهای کلان نظیر پاندمی)، ما نتیجه می‌گیریم که **تجميع داده‌های طولی و استفاده از مدل‌های جامع، قوی‌ترین استراتژی برای پیش‌بینی دقیق آینده** است. الگوهای پیچیده و پایدار شناسایی شده بر ضرورت سیاست‌گذاری شخصی‌سازی شده و مداخلات چندوجهی، با تمرکز بر گروه‌های سنی خاص و یک رویکرد یکپارچه برای حمایت از سلامت روان، سلامت جسم و عوامل اجتماعی-اقتصادی، تأکید می‌کند. این نتایج داده‌محور می‌توانند ابزاری حیاتی برای طراحی برنامه‌های مؤثرتر در بهداشت عمومی و کنترل دخانیات فراهم آورند.

۱. مقدمه

ترک موفق سیگار، به عنوان یکی از مؤلفه‌های کلیدی در ارتقاء بهداشت عمومی، پدیده‌ای چندوجهی است که تحقیقات بسیاری با استفاده از داده‌های نظرسنجی بزرگ مانند سیستم نظارت بر عوامل خطر رفتاری (BRFSS) به شناسایی عوامل مؤثر بر آن پرداخته‌اند. با این حال، بسیاری از این مطالعات بر پایه مدل‌های آماری خطی استوارند که ممکن است قادر به ثبت کامل روابط پیچیده و تعاملی بین متغیرها نباشند. به علاوه، پایداری زمانی این عوامل در طول یک دوره شش ساله (۲۰۱۸-۲۰۲۳) که شامل تحولات اجتماعی بزرگی مانند پاندمی کووید-۱۹ بوده، کمتر مورد بررسی عمیق قرار گرفته است. این تحقیق با هدف پر کردن این شکاف‌های روش‌شناختی و مفهومی، از یک رویکرد یادگیری ماشین پیشرفته بر روی داده‌های طولی از BRFSS استفاده می‌کند.

هدف اصلی ما، فراتر از ساخت یک مدل پیش‌بینی‌کننده با دقت بالا، دستیابی به یک درک عمیق، پایدار و قابل‌تفسیر از پویایی‌های ترک سیگار است. برای این منظور، این پژوهش به دنبال پاسخ به چندین سوال و فرضیه کلیدی است که در سه سطح تعریف می‌شوند:

۱. آزمون فرضیه‌های بنیادین (Validation of Foundational Hypotheses): در گام نخست، ما به بازآزمایی و

سنجش فرضیه‌هایی می‌پردازیم که بر اساس بدنه اصلی دانش موجود در این حوزه شکل گرفته‌اند:

- **فرضیه تاریخچه مصرف و سلامت فردی:** ما این فرضیه را می‌آزماییم که یک تاریخچه مصرف سبک‌تر (مانند سن شروع بالاتر و تعداد سیگار کمتر) و شاخص‌های سلامت جسمی بهتر (مانند BMI در محدوده نرمال) به طور معناداری شانس موفقیت در ترک سیگار را افزایش می‌دهند، در حالی که مشکلات سلامت روان به عنوان یک مانع مهم عمل می‌کند.

- **فرضیه حمایت اجتماعی و تعامل محیطی:** ما انتظار داریم که عوامل مرتبط با حمایت اجتماعی (به خصوص وضعیت تأهل) نقش مثبتی در افزایش شانس ترک داشته باشند و این تأثیر ممکن است توسط مشخصه‌های جغرافیایی (مانند محل سکونت شهری/روستایی) تعدیل شود.

۲. تبیین روابط پیچیده و متناقض (Explaining Complex and Paradoxical Relationships): یکی از چالش‌های اصلی در ادبیات علمی موجود، وجود نتایج متناقض در مورد عوامل اجتماعی-اقتصادی است. مدل‌های خطی سنتی اغلب یک رابطه مثبت و ساده بین عواملی مانند سن، درآمد و تحصیلات با ترک سیگار نشان می‌دهند. این تحقیق به دنبال پاسخ به این سوال است که: آیا مدل‌های یادگیری ماشین غیرخطی ما قادر به روشن‌سازی این تضادها هستند؟ مشخصاً، ما ماهیت واقعی رابطه این متغیرها را بررسی می‌کنیم تا مشخص شود آیا این عوامل، برخلاف تصور رایج، می‌توانند به عنوان مانع عمل کنند یا خیر.

۳. ارزیابی پایداری زمانی (Assessment of Temporal Stability): یک سوال محوری در تحلیل ما این است که آیا الگوهای پیش‌بینی‌کننده فوق، در طول دوره شش ساله مورد مطالعه و علی‌رغم تغییرات اجتماعی، ثابت باقی مانده‌اند؟ پاسخ به این سوال به ما کمک می‌کند تا بفهمیم آیا می‌توانیم به یک مدل جامع و قابل تعمیم برای آینده دست یابیم یا خیر.

با به کارگیری مدل‌هایی مانند LightGBM و روش تفسیرپذیری SHAP، این پژوهش نه تنها به شناسایی مهم‌ترین عوامل می‌پردازد، بلکه به تحلیل نحوه، جهت و پایداری زمانی تأثیرگذاری آن‌ها نیز می‌پردازد تا در نهایت، شواهدی داده‌محور برای طراحی مداخلات بهداشت عمومی هدفمندتر و شخصی‌سازی شده ارائه دهد.

2. روش‌شناسی (Methodology)

2.1. طراحی پژوهش و منبع داده

این پژوهش یک مطالعه کمی و تحلیلی با هدف اصلی شناسایی و ارزیابی عوامل پیش‌بینی‌کننده ترک موفق سیگار است. برای افزایش اعتبار و ارزیابی پایداری یافته‌ها، ما از یک طراحی تحلیل طولی با استفاده از مقاطع عرضی تکرارشونده (repeated cross-sectional) بهره بردیم. داده‌های مورد استفاده از نظرسنجی سیستم نظارت بر عوامل خطر رفتاری (BRFSS)، که توسط مرکز کنترل و پیشگیری از بیماری‌ها (CDC) در ایالات متحده مدیریت می‌شود، استخراج گردید. این تحلیل بر روی داده‌های سال‌های ۲۰۱۸ تا ۲۰۲۳ (با در نظر گرفتن سال‌های موجود) متمرکز بود تا اطمینان حاصل شود که عوامل شناسایی شده در طول یک دوره زمانی پرتلاطم، پایدار هستند.

2.2. آماده‌سازی داده و تعریف جامعه نمونه

فرآیند آماده‌سازی داده‌ها با استخراج مجموعه‌ای از ویژگی‌های بالقوه مؤثر از کدبک‌های BRFSS آغاز شد که حوزه‌های

دموگرافیک، اجتماعی-اقتصادی، سلامتی و رفتاری را پوشش می‌دادند. ویژگی‌هایی که کیفیت داده پایینی داشتند (مثلاً درصد مقادیر گم‌شده بالا)، از تحلیل حذف شدند. جامعه نمونه ما شامل پاسخ‌دهندگانی بود که سابقه مصرف حداقل ۱۰۰ نخ سیگار را در طول عمر خود گزارش کرده بودند. ($\text{smoke100}=1$)

متغیر هدف باینری، **ترک موفق سیگار (کلاس ۱)**، از روی متغیر محاسبه شده **SMOKER3** برای "سیگاری‌های سابق" در مقابل **ادامه مصرف (کلاس ۰)** برای "سیگاری‌های فعلی" تعریف شد. رکوردهای نامرتبط (مانند افراد غیرسیگاری یا پاسخ‌های نامشخص) حذف گردیدند. همچنین، برای جلوگیری از **نشت اطلاعات (Data Leakage)**، ویژگی‌هایی که ذاتاً با متغیر هدف هم‌پوشانی داشتند (مانند x_{rfsmok3} , x_{smoke100}) از مجموعه متغیرهای پیش‌بینی‌کننده حذف شدند.

2.3. استراتژی‌های انتخاب و تحلیل ویژگی‌ها

برای شناسایی مهم‌ترین عوامل از میان مجموعه گسترده‌ای از متغیرها، چندین استراتژی انتخاب ویژگی به کار گرفته و نتایج آن‌ها با یکدیگر مقایسه شد:

- **روش‌های آماری: (Univariate)** مانند SelectKBest برای ارزیابی قدرت تفکیک هر ویژگی به صورت مجزا.
 - **روش‌های مبتنی بر مدل: (Multivariate)** مانند $\text{RandomForest Feature Importance}$ و $\text{Recursive Feature Elimination (RFE)}$ که تعاملات بین ویژگی‌ها را در نظر می‌گیرند.
 - **کاهش هم‌خطی با VIF:** یک سناریوی تحلیلی مجزا نیز برای ارزیابی تأثیر هم‌خطی بر مدل‌ها، با استفاده از زیرمجموعه‌ای از ویژگی‌ها با $VIF < 7$ ، طراحی شد.
- پس از مقایسه اولیه، تصمیم گرفته شد که مدل‌های نهایی روی دو سناریو اصلی اجرا شوند: ۱) **استفاده از تمام ویژگی‌های باقی‌مانده** (پس از پاک‌سازی اولیه) و ۲) **استفاده از ویژگی‌های منتخب $VIF > 7$** تا تأثیر هم‌خطی بر عملکرد مدل‌ها به صورت تجربی ارزیابی شود.

2.4. فرآیند مدل‌سازی و ارزیابی

برای دستیابی به یک دید جامع، از طیف وسیعی از الگوریتم‌های یادگیری ماشین استفاده شد:

- **مدل‌های کلاسیک و قابل تفسیر:** رگرسیون لجستیک، درخت تصمیم، و تحلیل تشخیصی خطی. (LDA)
 - **مدل‌های آنسامبل بوستینگ گرادیان:** XGBoost, LightGBM, و CatBoost.
- تمام مراحل پیش‌پردازش ($\text{imputation, scaling, encoding}$) و مدل‌سازی در یک چارچوب Pipeline استاندارد پیاده‌سازی شد. **وزن‌های نظرسنجی (LLCPWT)** برای تعمیم‌پذیری بهتر نتایج و پارامتر scale_pos_weight برای

مدیریت عدم توازن کلاس‌ها در فرآیند آموزش مدل‌ها لحاظ گردید. عملکرد مدل‌ها با استفاده از معیارهای استاندارد شامل **F1-Score، AUC-ROC، و دقت (Accuracy)** ارزیابی شد.

2.5. ساخت و ارزیابی مدل جامع

برای ساخت یک مدل نهایی و قدرتمند که از دانش تمام سال‌ها بهره‌برد و قادر به پیش‌بینی روی داده‌های جدید باشد، دو رویکرد پیشرفته مدل‌سازی جامع به کار گرفته و با هم مقایسه شدند:

1. **آنسامبل زمانی: (Temporal Ensemble)** مدل‌های LightGBM که برای هر سال به صورت جداگانه آموزش داده شده بودند، در یک ساختار آنسامبل با هم ترکیب شدند. پیش‌بینی نهایی از طریق میانگین ساده و میانگین وزن‌دار (با وزن‌های بهینه‌سازی شده) از احتمالات خروجی هر مدل به دست آمد.
2. **تجمیع داده: (Data Pooling)** تمام داده‌های سال‌های موجود در یک دیتافریم واحد تجمیع شدند و یک ویژگی **year** به عنوان متغیر کنترلی به مدل اضافه شد. سپس مدل‌های برتر (مانند LightGBM و CatBoost) بر روی این مجموعه داده جامع آموزش داده شدند تا یک مدل واحد و قدرتمند ساخته شود.
3. **روش انباشت (Stacking Ensemble):** علاوه بر مدل‌های جامع بالا، برای ارتقای بیشتر عملکرد و بهره‌گیری از نقاط قوت مدل‌های مختلف، روش انباشت (Stacking Ensemble) نیز پیاده‌سازی و ارزیابی شد. در این رویکرد، مجموعه‌ای از پایپ‌لاین‌های مدل‌های برتر که شامل بهترین پیکربندی‌های LightGBM و XGBoost به عنوان مدل‌های پایه (Base Models) استفاده شدند. خروجی این مدل‌های پایه، به عنوان ورودی برای یک «فرامدل» (Meta-Model) رگرسیون لجستیک (Logistic Regression) در نظر گرفته شد. فرآیند آموزش با استفاده از اعتبارسنجی متقابل (Cross-Validation با $cv=5$) انجام گردید تا از صحت نتایج اطمینان حاصل شود. هدف این رویکرد، یادگیری پیچیدگی‌ها از طریق مدل‌های پایه و سپس ترکیب بهینه این یادگیری‌ها توسط فرامدل برای تولید پیش‌بینی نهایی و بهبود عملکرد کلی مدل بود.

2.6. تحلیل تفسیرپذیری (Model Interpretability)

برای پاسخ به سوالات تحقیق و درک عمیق "چگونگی" تأثیرگذاری هر عامل بر ترک سیگار، ما از یک رویکرد دوگانه برای تفسیرپذیری مدل‌ها استفاده کردیم:

1. **تحلیل ضرایب در مدل‌های خطی:** برای مدل‌های قابل تفسیر مانند رگرسیون لجستیک و تحلیل تشخیصی خطی (LDA)، ما به تحلیل مستقیم ضرایب (Coefficients) هر ویژگی پرداختیم. برای رگرسیون لجستیک، این ضرایب به نسبت شانس (Odds Ratios) نیز تبدیل شدند تا تفسیر آن‌ها ساده‌تر شود. این روش به ما اجازه داد تا اثر خالص و جهت هر متغیر را با فرض یک رابطه خطی و با کنترل سایر ویژگی‌ها، ارزیابی کنیم.

۲. تحلیل مبتنی بر نظریه بازی‌ها برای مدل‌های بوستینگ: برای باز کردن "جعبه سیاه" مدل‌های بوستینگ پیچیده (LightGBM, CatBoost, XGBoost)، از روش پیشرفته **SHAP (SHapley Additive exPlanations)** استفاده شد. مقادیر SHAP، که بر اساس نظریه بازی‌های مشارکتی محاسبه می‌شوند، به ما اجازه دادند تا سهم و مشارکت هر ویژگی را در پیش‌بینی هر نمونه به صورت جداگانه تعیین کنیم. با استفاده از نمودارهای خلاصه‌ی SHAP، ما توانستیم نه تنها اهمیت کلی هر ویژگی، بلکه جهت و اندازه تأثیر آن را نیز در سطح کل مجموعه داده شناسایی کرده و پایداری زمانی این الگوهای غیرخطی را در طول سال‌های مختلف ارزیابی کنیم.

4. یافته‌ها (results):

4.1. اطلاعات کلی درباره ویژگی‌ها

ثبات نسبت کلاس هدف: در طول سال‌های مورد مطالعه (۲۰۱۸-۲۰۲۳)، نسبت افراد در کلاس "ترک موفق سیگار" (کلاس 1) در مقایسه با کلاس "ادامه مصرف سیگار" (کلاس 0) نسبتاً ثابت باقی مانده است. درصد کلاس 1 در حدود ۶۶٪ تا ۷۱٪ نوسان داشته (مثلاً ۶۵.۹۶٪ در 2018 تا ۷۱.۴۷٪ در 2023). این بدان معنی است که در تمام این سال‌ها، جمعیت "سیگاری سابق" (Former Smoker) بخش عمده‌ای از نمونه مورد مطالعه را تشکیل داده است.

ویژگی‌های انتخاب‌شده بر اساس معیار $VIF > 7$ (با هدف کاهش هم‌خطی) به شرح زیر هستند: ['x_age80', 'x_bmi5', 'weight2', 'wtkg3', 'height3', 'children', 'hhadult', 'alcdays5', 'menthlth', 'physlth', 'poorhlth', 'x_casthm1', 'x_totinda', 'x_ment14d', 'x_phys14d', 'x_imprace', 'x_prace1', 'marital', 'x_incomg', 'employ1', 'x_state', 'persdoc2', 'checkup1', 'x_rfhth']

4.2. عملکرد مدل‌های پیش‌بینی‌کننده

عملکرد کلی مدل‌ها در پیش‌بینی ترک سیگار در طول سال‌های ۲۰۱۸ تا ۲۰۲۳ از روندهای زیر تبعیت کرده است:

- **برتری پایدار مدل‌های بوستینگ:** در تمامی سال‌ها، مدل‌های مبتنی بر گرادیان بوستینگ (به‌ویژه LightGBM و CatBoost) در مقایسه با مدل‌های کلاسیک مانند رگرسیون لجستیک، درخت تصمیم و LDA، به طور مداوم بالاترین مقادیر AUC، Accuracy و F1-Score را کسب کرده‌اند.
- **LightGBM (All Features) و CatBoost (All Features)** همواره در رتبه‌های اول یا دوم از نظر AUC (محدوده ۰.۷۷ تا ۰.۷۹) و F1-Score (محدوده ۰.۷۱ تا ۰.۷۲) قرار گرفته‌اند.
- این روند ثابت، توانایی بالای این مدل‌های غیرخطی را در پیش‌بینی ترک سیگار از داده‌های BRFSS نشان می‌دهد.

- **مزیت حفظ "تمام ویژگی‌ها" (All Features):** در تمام سال‌ها، سناریوی مدل‌سازی با استفاده از تمام ویژگی‌های پس از پیش‌پردازش (All Features)، به طور مداوم عملکرد بهتری را (با اختلافی حدود ۱ تا ۲ درصد در AUC) نسبت به سناریوی ویژگی‌های منتخب با (معیار $VIF > 7$) به دست آورده است.

• مدل‌های کلاسیک:

- رگرسیون لجستیک (All Features) و LDA (All Features) نیز عملکرد قابل قبولی از خود نشان داده‌اند، با مقادیر AUC که معمولاً بین ۰.۷۴ تا ۰.۷۷ نوسان داشتند. با این حال، آنها آنها به سطح عملکرد مدل‌های بوستینگ نرسیده‌اند.
- در سال ۲۰۲۳، مدل‌های LDA (All Features) و Logistic Regression (All Features) به طور غیرمنتظره‌ای بالاترین دقت (Accuracy) کلی را در جدول مقایسه عملکرد نشان دادند (LDA با 0.7390 و Logistic Regression با 0.7373). این نوسان در دقت کلی، علی‌رغم AUC های پایین‌تر نسبت به مدل‌های بوستینگ، مشاهده شد.

- **مدل PCA + Logistic Regression:** در تمام سال‌ها (۲۰۱۸-۲۰۲۳)، مدل Logistic Regression پس از کاهش ابعاد با PCA، به طور مداوم دچار مشکل "پیش‌بینی همه به سمت کلاس مثبت (ترک موفق)" بود. این مدل Precision, Recall و F1-Score نزدیک به صفر برای کلاس "ادامه مصرف سیگار" (کلاس 0) داشت (معمولاً 0.00 در Recall برای کلاس 0). این الگو نشان می‌دهد که رویکرد PCA (با حفظ ۹۵٪ واریانس و انتخاب تنها ۳ یا ۴ مؤلفه اصلی) برای پیش‌بینی عدم موفقیت در ترک سیگار ناکارآمد است. این مشکل در طول تمامی سال‌های تحلیل شده تکرار شد.

4.3. عملکرد مدل‌های جامع (Ensemble and Pooled Data Models)

برای بهره‌گیری حداکثری از داده‌های موجود و بهبود قدرت تعمیم‌پذیری مدل در طول زمان و بر روی داده‌های جدید، دو رویکرد جامع "انسامبل زمانی" (Temporal Ensemble) و "تجمیع داده" (Data Pooling)، به همراه یک الگوی پیشرفته "روش انباشت" (Stacking Ensemble) پیاده‌سازی و ارزیابی شدند.

الف. روش تجمیع داده (Data Pooling):

مدل‌های LightGBM و CatBoost که بر روی کل داده‌های تجمیع‌شده (۲۰۱۸-۲۰۲۳) و با افزودن متغیر year آموزش داده شدند، عملکرد بالایی از خود نشان دادند:

| Model | Accuracy | F1-Score | AUC |
|-------------------------|----------|----------|----------|
| CatBoost (All Features) | 0.742545 | 0.729380 | 0.783549 |
| LightGBM (All Features) | 0.742213 | 0.728049 | 0.782297 |
| XGBoost (All Features) | 0.740577 | 0.727649 | 0.781083 |

- **تحلیل ماتریس پیچیدگی (Confusion Matrix):** در این مدل‌ها، ماتریس‌های پیچیدگی همچنان الگوی غالب بودن **True Positives** نسبت به **True Negatives** را حفظ کردند که نشان‌دهنده توانایی بالا در پیش‌بینی ترک موفق، اما با سوگیری به سمت کلاس اکثریت بود. منحنی‌های ROC و Precision-Recall نیز با AUC 0.78 و AP 0.88، عملکرد بالایی کلی مدل را تأیید کردند.
- **تحلیل اهمیت ویژگی‌ها (Feature Importance)** در این مدل‌های تجمیع‌شده نشان داد که ویژگی‌های اصلی همانند تحلیل‌های سالانه پایدار هستند، با x_age5yr (سن)، marital (وضعیت تأهل)، و $chccopd1$ (COPD) در بالاترین رتبه‌های اهمیت قرار داشتند. همچنین، متغیر $year$ نیز با اهمیت قابل قبول ۰.۱۱۴۹، نقش زمان را در پیش‌بینی‌ها نشان داد.
- **تحلیل SHAP در مدل‌های تجمیع‌شده** نیز پایداری الگوهای تأثیرگذاری ویژگی‌ها را تأیید کرد:
 - **سن (x_age80, x_age5yr):** بالا بودن سن با افزایش احتمال ترک سیگار همراه بود.
 - **وضعیت تأهل (marital):** افراد متأهل (آبی) شانس ترک را افزایش و سایر وضعیت‌ها شانس ترک را کاهش می‌دادند.
 - **درآمد (x_incomg):** درآمد بالاتر، افزایش احتمال ترک را نشان داد.
 - **BMI (x_bmi5) و وزن ($weight2, wtkg3$):** BMI بالاتر و وزن بیشتر (قرمز) افزایش احتمال ترک سیگار را به همراه داشتند.
 - **بیماری‌های جدی ($chccopd1, addepev2$):** سابقه بیماری (آبی) افزایش شانس ترک سیگار را نشان داد.
 - **سلامت عمومی (genhlth):** سلامت بدتر کاهش شانس ترک را به همراه داشت.
 - **سلامت روان نامناسب ($x_ment14d$):** مشکلات سلامت روان به طور پیوسته کاهش احتمال ترک را نشان داد.
 - **فعالیت فیزیکی ($x_totinda$):** فعالیت فیزیکی در مدل CatBoost به طور ثابت افزایش ترک سیگار و عدم فعالیت فیزیکی کاهش احتمال ترک سیگار دیده می‌شود، که نشان‌دهنده ارتباط مثبت فعالیت با ترک سیگار است.
 - **تحصیلات (x_educag):** تحصیلات بالاتر افزایش احتمال ترک سیگار را نشان داد.
 - **آخرین چکاپ پزشکی ($checkup1$):** چکاپ‌های قدیمی‌تر یا عدم انجام چکاپ کاهش احتمال ترک سیگار را نشان داد.

ب. روش آنسامبل زمانی (Temporal Ensemble):

این رویکرد که خروجی‌های مدل‌های LightGBM هر سال را ترکیب می‌کند، عملکردی قوی و پایدار را نشان داد:

| F1-Score | Accuracy | AUC | Method |
|----------|----------|----------|-----------------------------|
| 0.731618 | 0.743469 | 0.785189 | Simple Average Ensemble |
| 0.729952 | 0.743481 | 0.785549 | Weighted Average Ensemble |
| 0.731618 | 0.743469 | 0.785189 | Optimized Weighted Ensemble |

نزدیکی عملکرد بین روش‌های مختلف وزن‌دهی (ساده، وزن‌دار ثابت و وزن‌دار بهینه‌شده با وزن‌های مساوی) نشان‌دهنده پایداری بالای عملکرد مدل LightGBM در طول هر سال و همسانی قدرت پیش‌بینی آن در دوره‌های زمانی مختلف است. بهترین وزن‌های بهینه‌سازی شده [0.167 0.167 0.167 0.167 0.167 0.167] نیز گواهی بر همین همسانی وزن است.

ج. روش انباشت (Stacking Ensemble):

رویکرد Stacking، با ترکیب بهترین مدل‌های پایه (LightGBM و XGBoost) توسط یک متا-مدل رگرسیون لجستیک، بالاترین عملکرد کلی را در میان تمامی رویکردهای مدل‌سازی (از جمله مدل‌های انفرادی و آنسامبل زمانی) به دست آورد.

نتایج ارزیابی Stacking Ensemble: AUC: 0.7881 , Accuracy: 0.7446 , F1: 0.7330

ضرایب متا-مدل: [2.49886214:(LightGBM), 2.66476161:(XGBoost)]

ضرایب مثبت متا-مدل رگرسیون لجستیک برای LightGBM و XGBoost نشان می‌دهد که پیش‌بینی‌های هر دو مدل پایه تأثیر مثبت و قابل توجهی بر پیش‌بینی نهایی متا-مدل داشته‌اند. مقدار کمی بالاتر برای XGBoost نشان‌دهنده اهمیت کمی بیشتر (اما همچنان مشابه) آن در ترکیب نهایی است.

خلاصه: این نتایج نشان‌دهنده توانایی Stacking در بهره‌برداری از نقاط قوت مدل‌های پایه برای رسیدن به بالاترین سطح دقت پیش‌بینی ممکن در این مجموعه داده است، و این مدل به عنوان قوی‌ترین مدل پیش‌بین ترک سیگار در این پژوهش مطرح می‌گردد.

4.4. بررسی ماتریس‌های پیچیدگی (ConfusionMatrix)

سوگیری ثابت به سمت کلاس اکثریت: در تمامی مدل‌های بوستینگ در هر دو سناریوی VIF و All Features، یک الگوی ثابت در ماتریس‌های صحت مشاهده شد:

- تعداد **True Positives** (پیش‌بینی صحیح ترک موفق) به طور مداوم و قابل توجهی بسیار بیشتر از تعداد **True Negatives** (پیش‌بینی صحیح ادامه مصرف) بود (غالباً True Positives تقریباً دو برابر True Negatives بود).

4.5. تحلیل اهمیت ویژگی‌ها و تأثیر آن‌ها در مدل‌های گرادیان بوستینگ (SHAP values)

نکته در تحلیل SHAP: مقادیر SHAP مثبت به معنی افزایش احتمال ترک موفق سیگار، و مقادیر منفی SHAP به معنای کاهش احتمال ترک موفق سیگار هستند.

در تحلیل مقادیر SHAP (برای مدل‌های بوستینگ) که ابزاری قدرتمند برای تفسیر تأثیر هر ویژگی در مدل‌های غیرخطی است، روندهای پایداری در طول سال‌های ۲۰۱۸ تا ۲۰۲۳ مشاهده شد:

• تأثیر غالب و پایدار سن (x_{age_g} , x_{age5yr} , x_{age80}):

- روند و تغییر: این متغیرها به طور مداوم قوی‌ترین عامل‌های پیش‌بینی در تمام سال‌ها بودند. سن بالاتر (نقاط قرمز)، به طور قطع و مداوم، با افزایش احتمال ترک موفق سیگار مرتبط بودند.
- تأثیر پایدار وضعیت تأهل (**marital**): این ویژگی به طور مداوم در بین پنج ویژگی برتر از نظر اهمیت کلی در تمام سال‌ها قرار داشت.

- در نمودارهای SHAP، به وضوح مشاهده می‌شود که نقاط آبی رنگ ('Married' / متأهل) به طور پیوسته و در تمام سال‌های مورد مطالعه، احتمال ترک موفق سیگار را در پیش‌بینی مدل افزایش می‌دهد.

- در مقابل، نقاط قرمز رنگ (وضعیت‌های "A", "Never married", "Separated", "Widowed", "Divorced", "member of an unmarried couple") که سایر وضعیت‌های تأهل (غیر از متأهل)، به طور پیوسته و در تمام سال‌ها، احتمال ترک موفق سیگار را در پیش‌بینی مدل کاهش می‌دهند.

- تأثیر درآمد (x_{incomg}): درآمد بالاتر (نقاط قرمز) به طور پایدار و باثبات، با افزایش احتمال ترک موفق سیگار همراه بوده‌اند. این نتیجه نیز یک روند ثابت و مهم در تمام سال‌ها است.

- تأثیر BMI (x_{bmi5}): مقادیر بالاتر BMI (نقاط قرمز) به طور مداوم و باثبات، با افزایش احتمال ترک موفق سیگار همراه بوده‌اند.

• روند سلامت روان و جسم:

- تشخیص بیماری‌های حاد ($addepev2$, $cvdstrk3$, $chccopd1$): در تمامی سال‌ها، تشخیص بیماری‌های حاد (مانند COPD، سکته مغزی، یا افسردگی بالینی) به طور قاطع با افزایش احتمال "ترک موفق سیگار" در ارتباط بوده است.

- نقاط آبی رنگ ("سابقه COPD دارد") ، داشتن سابقه بیماری مزمن انسدادی ریه" به طور پیوسته، احتمال ترک موفق سیگار را در پیش‌بینی مدل افزایش می‌دهد. در مقابل، نقاط قرمز رنگ (سابقه COPD ندارد") به کاهش احتمال ترک سیگار اشاره دارند.

- داشتن سابقه سکته مغزی به طور پیوسته، احتمال ترک موفق سیگار را در پیش‌بینی مدل افزایش می‌دهد. در مقابل، نداشتن سابقه سکته مغزی باعث کاهش احتمال ترک سیگار می‌شود.

- همچنین "داشتن سابقه تشخیص اختلال افسردگی"، احتمال ترک موفق سیگار را در پیش‌بینی مدل افزایش می‌دهد. در مقابل، "نداشتن سابقه اختلال افسردگی" احتمال ترک موفق سیگار را در پیش‌بینی مدل کاهش می‌دهد.

- سلامت عمومی (genhlth): در این سالها، نقاط قرمز رنگ (نشان‌دهنده وضعیت سلامت بدتر / 'Fair', 'Poor') عمدتاً احتمال "ترک موفق سیگار" کمتری را در پیش‌بینی مدل داشته‌اند. برعکس، نقاط آبی (سلامت بهتر) اغلب احتمال ترک موفق سیگار بیشتری را در پیش‌بینی مدل داشته‌اند.

- سلامت روان نامناسب (x_ment14d ,menthlth): در تمامی سال‌ها، به ویژه از ۲۰۲۰/۲۰۱۹ به بعد که الگو تثبیت شد، "مشکلات سلامت روان (نظیر روزهای بیشتر با روحیه پایین)"، به طور پایداری، احتمال ترک موفق سیگار را در پیش‌بینی مدل کاهش می‌دهد. برعکس، "سلامت روان بهتر" با تمایل بیشتری برای ترک موفق سیگار همراه است.

• سایر ویژگی‌ها:

- x_totinda (فعالیت فیزیکی یا ورزشی): در تمامی سال‌ها و در هر دو سناریوی All Features و VIF برای مدل CatBoost، یک الگوی کاملاً ثابت و واضح مشاهده می‌شود.

- نقاط آبی رنگ ("داشتن فعالیت فیزیکی"=1) عمدتاً احتمال "ترک موفق سیگار" را در پیش‌بینی مدل افزایش می‌دهد. نقاط قرمز رنگ ("عدم فعالیت فیزیکی"=2) در بیشتر مواقع احتمال "ترک موفق سیگار" را در پیش‌بینی مدل کاهش می‌دهد.

- به وضوح نشان می‌دهد که "فعالیت فیزیکی"، عامل "افزایش‌دهنده" احتمال ترک سیگار، و "عدم فعالیت فیزیکی" عامل "کاهش‌دهنده" آن است.

- x_educag: همواره در رده میانی تا بالای جدول اهمیت قرار داشته است که یعنی از اهمیت قابل قبولی برخوردار است. در تمامی سال‌ها و در تمامی مدل‌های بوستینگ (با All Features و VIF)، یک الگوی کاملاً ثابت و واضح مشاهده می‌شود:

- نقاط قرمز رنگ (یعنی کدهای 3 و 4 که به ترتیب "شرکت در کالج یا مدرسه فنی" و "فارغ التحصیلی از کالج یا مدرسه فنی") یعنی "داشتن سطوح تحصیلات بالاتر"، عمدتاً در تمام سال‌ها، احتمال "ترک موفق سیگار" را در پیش‌بینی مدل افزایش می‌دهد.
- نقاط آبی رنگ (مقادیر پایین این ویژگی، یعنی کدهای 1 و 2 که به ترتیب "عدم فارغ التحصیلی از دبیرستان" و "فارغ التحصیلی از دبیرستان") یعنی "داشتن سطوح تحصیلات پایین‌تر"، عمدتاً در تمام سال‌ها، احتمال "ترک موفق سیگار" را در پیش‌بینی مدل کاهش می‌دهد.
- آخرین معاینه پزشکی (1checkup): در تمامی سال‌ها، نقاط قرمز رنگ ("چکاپ‌های قدیمی‌تر یا عدم انجام چکاپ") به طور مداوم احتمال "ترک موفق سیگار" را در پیش‌بینی مدل کاهش می‌دهد. برعکس، نقاط آبی (چکاپ جدیدتر) یعنی داشتن چکاپ‌های منظم‌تر (دسترسی به پزشک و مشاوره سلامت) اغلب به شانس بالاتری برای ترک موفق سیگار همراه است.
- تاثیر وزن (num__weight2 و num__wtkg3): در تمامی سال‌ها و در بیشتر مدل‌های بوستینگ، به طور مداوم در رده‌های میانی SHAP قرار داشته‌اند. در تمامی سال‌ها (۲۰۱۸ تا ۲۰۲۳) و در تمام مدل‌های بوستینگ (LightGBM, CatBoost, XGBoost) و در هر دو سناریوی All Features و VIF، یک الگوی کاملاً ثابت، واضح و پیوسته مشاهده می‌شود:
- نقاط قرمز رنگ (مقادیر بالا برای وزن، یعنی افراد سنگین‌تر) به طور غالب در تمام طول دوره مطالعه، احتمال "ترک موفق سیگار" را در پیش‌بینی مدل افزایش می‌دهد. در مقابل، نقاط آبی رنگ (مقادیر پایین برای وزن، یعنی افراد سبک‌تر) عمدتاً احتمال "ترک موفق سیگار" را در پیش‌بینی مدل کاهش می‌دهد.
- x_state (ایالت): در برخی سال‌ها (مثلاً ۲۰۲۰)، تأثیر قابل توجهی در SHAP داشت که می‌تواند نشان‌دهنده واریانس جغرافیایی باشد.

4.6. تحلیل ضرایب رگرسیون لجستیک (Logistic Regression Coefficient Analysis)

- در تحلیل ضرایب رگرسیون لجستیک، که بیانگر رابطه خطی بین هر ویژگی و لگاریتم شانس ترک سیگار است، روابط و روندهای زیر مشاهده شد:
- **پایداری کلی:** بسیاری از ویژگی‌های کلیدی که در تحلیل SHAP برجسته بودند (به‌ویژه سن، وضعیت تأهل، درآمد و ویژگی‌های مرتبط با بیماری‌های جدی) دارای ضرایب و جهت‌گیری تأثیر پایداری در طول سال‌ها بودند.

- **تأثیر سن:** x_age80 همواره بالاترین ضریب مثبت را داشته است (با Odds Ratio های بین ۱.۴۷ تا ۲.۶۷). این تأییدکننده قوی ارتباط خطی مثبت بین سن بالا و شانس ترک سیگار در تمامی سال‌ها است.
- **سلامت روان:** ضرایب برای menthlth و x_ment14d اغلب کوچک، اما از سال 2020/2019 به بعد روندی افزایشی به سمت منفی شدن نشان دادند. به عنوان مثال، x_ment14d (روزهای با سلامت روان نامناسب) در سال‌های اخیر ضریب منفی و $OR > 1$ داشت، که به طور خطی نیز تأییدکننده نقش آن در ترک سیگار است. menthlth نیز همین الگو را در سال‌های اخیر نشان داد، اما با ضرایب نزدیکتر به صفر.
- **BMI (x_bmi5):** ضریب x_bmi5 نوسانات بیشتری نشان داد؛ در برخی سال‌ها (مثل VIF 2018، VIF 2023 و All Features) مثبت بود، و در برخی دیگر (مثلاً All Features 2020، All Features 2021) اندکی منفی.
- **فعالیت فیزیکی (x_totinda):** همواره دارای ضریب منفی قابل توجهی بود، که نشان‌دهنده کاهش شانس ترک سیگار با افزایش فعالیت فیزیکی است.
- **نقش بیماری‌ها:** متغیرهای مربوط به بیماری‌های جدی (مانند chccopd1، cvdstrk3، addepev2) همواره دارای ضرایب مثبت بالا بودند که تأیید کننده نقش این عوامل در ترک سیگار در یک رابطه خطی است.

4.7. قواعد درخت تصمیم‌گیری

- در نمودارهای درخت تصمیم‌گیری (چه در سناریو VIF و چه در All Features)، یک ساختار کلی و سلسله‌مراتب تصمیم‌گیری پایدار در طول سال‌ها مشاهده شد:
- **پایداری ویژگی‌های کلیدی در گره‌های اولیه:** x_age80 (سن)، marital (وضعیت تأهل)، x_incomg (درآمد)، و x_educag (سطح تحصیلات) به طور مداوم و در تمامی سال‌ها (۲۰۱۸-۲۰۲۳) در **گره‌های سطح بالای درختان** ظاهر شدند. این نشان‌دهنده نقش بنیادی و بسیار اولیه این ویژگی‌ها در تقسیم‌بندی نمونه‌ها و تصمیم‌گیری‌های اولیه برای پیش‌بینی ترک سیگار است، به این معنی که این ویژگی‌ها از **قوی‌ترین تفکیک‌کننده‌ها** هستند.
 - **ثبات ساختار درختی:** ساختار کلی و "شاخه‌های" اصلی درختان تصمیم در طول سالیان نسبتاً ثابت باقی مانده، هرچند جزئیات گره‌های پایین‌تر ممکن بود کمی متفاوت باشند. این پایداری ساختار، منعکس‌کننده پایداری الگوهای تصمیم‌گیری و اهمیت متغیرهای اصلی در پیش‌بینی ترک سیگار در طول دوره مطالعه است.

4.8. ضرایب همبستگی با هدف (x_smoker3)

در طول سال‌های مورد مطالعه (۲۰۱۸ تا ۲۰۲۳)، یک **پایداری قابل توجه** در جهت (مثبت یا منفی) و شدت همبستگی خطی میان ویژگی‌های مختلف و متغیر هدف x_smoker3 مشاهده شد.

- **x_llcpwt** (وزن نمونه‌گیری) و **smoke100** (سابقه مصرف ۱۰۰ نخ سیگار در زندگی): این دو ویژگی همواره بالاترین همبستگی مثبت را با متغیر هدف **x_smoker3** نشان دادند (معمولاً بالای ۰.۸).
- **x_rfsmok3** (سابقه سیگار کشیدن کنونی یا مصرف روزانه/غیرروزانه): این ویژگی به طور مداوم قوی‌ترین همبستگی منفی را با **x_smoker3** نشان داد (معمولاً بین -۰.۸ تا -۰.۹). این همبستگی قوی منفی نیز منطقی است، زیرا **x_rfsmok3** به احتمال زیاد متغیر مربوط به وضعیت "سیگاری فعال" را منعکس می‌کند، در حالی که **x_smoker3** وضعیت "سیگاری سابق" (ترک موفق) را نشان می‌دهد؛ لذا رابطه آن‌ها معکوس است.

2. ویژگی‌های با همبستگی ضعیف تا متوسط و جهت ثابت:

- اکثر قریب به اتفاق ویژگی‌های دموگرافیک، سلامت، و رفتاری دیگر، همبستگی‌های خطی نسبتاً ضعیفی (بین ۰ تا ۰.۲۵ یا ۰ تا -۰.۲۵) با متغیر هدف **x_smoker3** نشان دادند. مهمترین روند در این دسته، **ثبات جهت (مثبت یا منفی) همبستگی و نیز اندازه آن در طول تمامی سال‌های مشاهده بود:**
- **همبستگی‌های مثبت ثابت (نشان‌دهنده احتمال بیشتر "ترک سیگار موفق"):**

- ویژگی‌های مرتبط با سن (**x_ageg5yr, x_age80, x_age_g**): همبستگی مثبت (اغلب بین ۰.۰۵ تا ۰.۱۵). سن بالاتر، با احتمال بیشتری در ترک موفق سیگار همراه بود.
- **BMI (x_bmi5cat, x_bmi5)**: همبستگی مثبت (اغلب بین ۰.۰۵ تا ۰.۱). BMI بالاتر (یا در رده‌های بالاتر)، با احتمال بیشتری در ترک موفق سیگار مرتبط بود.
- **درآمد (x_incomg)**: همبستگی مثبت (اغلب بین ۰.۱ تا ۰.۲). درآمد بالاتر، احتمال ترک موفق سیگار بیشتری دارد.
- **تحصیلات (x_educag, educa)**: همبستگی مثبت (اغلب بین ۰.۰۵ تا ۰.۱۵). سطح تحصیلات بالاتر، با احتمال بیشتری در ترک موفق سیگار مرتبط بود.
- **مشکلات سلامتی/تشخیصی**
- **(x_asthms1, diffwalk, diabete3, addepev2, chcccpd1)**: همبستگی‌های مثبت ضعیف تا متوسط (معمولاً بین ۰.۰۵ تا ۰.۲۵) نشان می‌داد که تشخیص این بیماری‌ها یا مشکلات، با احتمال بیشتری در ترک موفق سیگار همراه بوده است.
- **سلامت روان و جسمی نامناسب (physhlth, menthlth)**: همبستگی مثبت ضعیف (اغلب بین ۰.۰۵ تا ۰.۱۰). تعداد روزهای بیشتر با سلامت روان/جسمی نامناسب، با احتمال بیشتری در ترک موفق همراه بود.

▪ همبستگی‌های منفی ثابت (نشان‌دهنده احتمال کمتر "ترک سیگار موفق"):

- **سلامت عمومی (genhlth):** همبستگی منفی (اغلب بین ۰.۱- تا ۰.۲۰-). سلامت عمومی بهتر، با احتمال کمتر ترک سیگار مرتبط بود.
- **فعالیت‌های بدنی (x_totinda):** همبستگی منفی ضعیف (اغلب بین ۰.۰۵- تا ۰.۱۵-). افراد با فعالیت بدنی جامع‌تر، با احتمال کمتری از ترک سیگار مرتبط بودند.
- **وزن بدنی و شاخص‌های مربوطه (children , wtkg3 ,weight2 , hhadult):** همبستگی‌های منفی بسیار ضعیف یا نزدیک به صفر.

4.9. SelectKBest با f_classif :

روش SelectKBest که بر پایه امتیاز f_classif (آزمون آنالیز واریانس برای طبقه‌بندی) ویژگی‌ها را انتخاب می‌کند، یک رویکرد **Univariate** است؛ یعنی قدرت تفکیک و تمایز هر ویژگی را به‌تنهایی و بدون در نظر گرفتن تعاملات یا هم‌خطی با سایر ویژگی‌ها، برای متغیر هدف ارزیابی می‌کند. در طول سال‌های مورد بررسی (۲۰۱۸ تا ۲۰۲۳)، روند زیر در این روش مشاهده می‌شود:

1. پایداری چشمگیر در ویژگی‌های برتر (ثبات رتبه‌بندی):

- **همواره در صدر:** در تمامی سال‌ها، سه ویژگی اصلی مرتبط با سن (یعنی x_age80، x_age5yr، و x_age_g) همواره و بدون استثناء، رتبه‌های اول، دوم و سوم را از نظر امتیاز f_classif به خود اختصاص داده‌اند. این نشان‌دهنده آن است که سن، به‌عنوان یک عامل منفرد، در تمامی مقاطع زمانی مورد مطالعه، قوی‌ترین قدرت تمایز را بین افراد "سیگاری سابق" و "سیگاری فعلی" داشته است.
- **ویژگی‌های پراهمیت ثابت:** پس از سن، مجموعه‌ای از ویژگی‌های مرتبط با **دموگرافیک، سلامت و سبک زندگی** نیز به طور مداوم در بین ۱۰ تا ۱۵ ویژگی برتر قرار داشته‌اند. این ویژگی‌ها عبارتند از:
 - **وضعیت تأهل (marital):** به‌طور پیوسته در رده‌های بالای اهمیت قرار دارد (معمولاً رتبه ۴ تا ۷).
 - **تحصیلات (x_educag, educa):** همواره جزء ویژگی‌های بسیار مهم بوده (معمولاً رتبه ۴ تا ۸).
 - **دسترسی به خدمات درمانی (persdoc2, checkup1, pneuvac4):** این ویژگی‌ها نیز به‌طور مداوم در رتبه‌های میانی تا بالای جدول اهمیت f_classif ظاهر شده‌اند.

- سلامت عمومی و روانی (x_ment14d, menthlth, addepev2, genhlth): ویژگی‌های مرتبط با وضعیت سلامت کلی و سلامت روان نیز پایداری خوبی در رتبه‌های بالای اهمیت از خود نشان داده‌اند.

- درآمد (x_incomg): در اکثر سال‌ها در رده ۱۰ تا ۱۵ ویژگی برتر قرار داشته است.

- BMI: نیز به طور مداوم در میان ۳۰ ویژگی برتر ظاهر شده‌اند.

2. کاهش جزئی در مقادیر مطلق امتیاز f_classif:

- روند: اگرچه رتبه‌بندی ویژگی‌ها در طول زمان بسیار ثابت باقی مانده، اما مقادیر مطلق

امتیاز f_classif برای ویژگی‌های برتر (مانند x_age80) از سال ۲۰۱۸ تا ۲۰۲۳، یک روند کلی کاهشی را

نشان می‌دهد. به عنوان مثال، امتیاز x_age80 از ۱۸۷۵۸ در سال ۲۰۱۸ به ۷۷۶۳ در سال ۲۰۲۳ کاهش یافته است.

4.10. RandomForest Feature Importance :

اهمیت ویژگی‌های RandomForest بر اساس میزان کاهش ناخالصی (impurity) توسط هر ویژگی در مجموعه‌ی درختان مدل ارزیابی می‌شود و به صورت ضمنی، تأثیر تعاملات بین ویژگی‌ها را نیز در نظر می‌گیرد. در طول سال‌های مورد بررسی (۲۰۱۸ تا ۲۰۲۳)، روندهای زیر در این معیار اهمیت مشاهده شد:

1. ثبات بی‌وقفه در مجموعه‌ی ویژگی‌های مهم:

- همواره در صدر: در تمامی سال‌ها، مجموعه‌ای ثابت از ۵ ویژگی اصلی، به طور مداوم بالاترین امتیاز اهمیت را در RandomForest به خود اختصاص داده‌اند:

- سن (x_age80) و گروه‌های سنی (x_age_g/x_ageg5yr): x_age80 به طور

غالب مهم‌ترین ویژگی یا در میان دو ویژگی اول از نظر اهمیت قرار داشته است (مثلاً 0.074 در

2018 تا 0.0658 در 2023). این پایداری نشان می‌دهد که سن قوی‌ترین و پایدارترین

پیشبینی ترک سیگار از منظر مدل‌های درختی است.

- BMI (x_bmi5): این ویژگی نیز همواره در میان سه یا چهار ویژگی اول قرار داشته است (اهمیت بین 0.064 تا 0.067).

- ایالت (x_state): در تمام سال‌ها، به طور قابل ملاحظه‌ای در رده‌های بالای اهمیت (معمولاً در

جایگاه سوم) قرار گرفته است (بین 0.054 تا 0.057). این ویژگی نسبت به روش f_classif

در RandomForest اهمیت بالاتری دارد.

○ وزن بدن و شاخص‌های وابسته (height3 , wtkg3 , weight2): این سه متغیر مرتبط با ویژگی‌های فیزیکی نیز به طور مداوم در بین ۱۰ ویژگی برتر قرار داشته‌اند (اهمیت معمولاً بین 0.045 تا 0.053).

- پایداری در میان ۱۰ تا ۲۰ ویژگی برتر: پس از هسته مرکزی ۵ ویژگی بالا، مجموعه‌ای دیگر از ویژگی‌ها شامل: درآمد (incomg)، وضعیت تأهل (marital)، تعداد روزهای مصرف الکل (alcdays)، وضعیت کلی سلامت (genhlth)، وضعیت استخدام (employ1) و آخرین چکاپ پزشکی (checkup1)، و همچنین وضعیت سلامت روان (physhlth , poorhlth , menthlth) به طور مداوم در میان ۲۰ ویژگی برتر اهمیت RandomForest در تمام سال‌ها مشاهده شده‌اند.

2. تغییرات جزئی در رتبه‌بندی صدرنشینان (به‌ویژه در سال‌های پایانی):

- برآمدگی BMI در سال ۲۰۲۳: یک تغییر قابل توجه در سال ۲۰۲۳ مشاهده شد، جایی که x_bmi5 با امتیاز ۰.۰۶۶۵، برای اولین بار بالاترین اهمیت را به دست آورد و جای x_age80 (با امتیاز ۰.۰۶۵۸) را در صدر گرفت.

5. مرور تحقیقات پیشین (Literature Review)

5.1. عوامل محیطی و سیاست‌گذاری در سطح ایالت

محیط سیاستی و فرهنگی که فرد در آن زندگی می‌کند، نقش مهمی در موفقیت ترک سیگار ایفا می‌کند. مطالعه ای مرتبط توسط ژو و همکاران (Zhu et al., 2025) در ژورنال JNCI این موضوع را به صورت مستقیم بررسی کرده است. آنها با استفاده از داده‌های BRFSS (2014-2019) و یک طراحی شبه‌آزمایشگاهی قوی، تأثیر افزایش عمده ۲ دلاری مالیات بر سیگار در کالیفرنیا (Proposition 56) را ارزیابی کردند. یافته‌های آنها نشان داد که این مداخله سیاستی بزرگ، همراه با سرمایه‌گذاری مجدد در برنامه‌های پیشگیری، به طور معناداری نرخ ترک سیگار را در کالیفرنیا (از ۱۱.۵٪ به ۱۴.۲٪) افزایش داد. در مقابل، ایالت‌هایی با افزایش مالیات کمتر یا بدون افزایش، هیچ بهبودی را تجربه نکردند. این یافته به ما در تفسیر یکی از نتایج کلیدی مدل‌هایمان کمک می‌کند، ویژگی x_state (ایالت) به طور مداوم در تحلیل اهمیت ویژگی با RandomForest در رتبه‌های بالا ظاهر می‌شود. این نشان می‌دهد که مدل‌های ما به صورت غیرمستقیم در حال یادگیری تأثیرات ناشی از تفاوت‌های سیاستی بین ایالت‌ها هستند.

در همین راستا، تحقیق جامع شاربا و همکاران (Sharbaugh et al., 2018) در PLOS ONE با استفاده از مدل‌های مختلط خطی (Linear mixed-effects models) بر روی داده‌های طولی BRFSS (۲۰۱۵-۲۰۰۱)، تأثیر مالیات‌های ایالتی بر شیوع سیگار کشیدن را بررسی کرد. آنها رابطه معکوس و معناداری بین این دو ویژگی یافتند: به ازای هر ۰.۲۵ دلار افزایش در مالیات،

شیوع مصرف به طور میانگین ۰.۶٪ کاهش می‌یافت. جالب توجه است که قوی‌ترین تأثیر در جوانان (۱۸-۲۴ سال) مشاهده شد (۱.۵٪ کاهش)، که نشان‌دهنده قدرت بازدارندگی قیمت برای شروع مصرف در این گروه است. همچنین، افزایش مالیات با افزایش تلاش برای ترک سیگار در میان سیگاری‌های فعلی مرتبط بود. با این حال، تأثیر بر افراد با **کمترین درآمد** ضعیف‌تر بود، که نویسندگان آن را به "رفتارهای کاهش هزینه" (مانند خرید برندهای ارزان‌تر) نسبت دادند.

جایگاه تحقیق حاضر در مقایسه با مطالعات فوق:

- **تمرکز بر ترک سیگار در مقابل شیوع:** در حالی که مطالعات ژو و شاربا بر تأثیر سیاست‌ها بر "شیوع" یا "تلاش برای ترک" تمرکز دارند، تحلیل ما بر "ترک موفق" در میان کسانی که قبلاً سیگاری بوده‌اند، متمرکز است. این دو معیار اگرچه مرتبط، اما جنبه‌های متفاوتی از مسئله را بررسی می‌کنند. تحقیق ما به طور خاص به فرآیند تغییر رفتار در افراد سیگاری می‌پردازد.
- **دامنه وسیع‌تر ویژگی‌ها و بروز زمانی:** ما مجموعه‌ای بسیار گسترده‌تر از ویژگی‌ها (شامل عوامل سلامت جسمی و روانی) را در مدل‌های خود گنجانده‌ایم و با پوشش دوره زمانی به‌روزتر (سال ۲۰۱۸ تا ۲۰۲۳)، پایداری این الگوها را در دوران بعد از ۲۰۱۹، از جمله دوران پاندمی، نشان می‌دهیم.

5.2. عوامل محیطی و نابرابری‌های جمعیتی در سطح ایالت

برای درک عمیق‌تر عوامل محیطی و جمعیتی، تحلیل‌های فراتر از سطح فردی ضروری است. اهمیت بالایی متغیرهایی مانند x_state (ایالت) و متغیرهای نژادی در مدل‌های پیشرفته ما، نشان می‌دهد که زمینه جغرافیایی-فرهنگی نقش مهمی ایفا می‌کند. در همین راستا، یک مطالعه کلیدی توسط میلز و همکاران (Mills et al., 2021) با استفاده از داده‌های BRFSS (۲۰۱۸-۲۰۱۱)، به بررسی ناهمگونی نابرابری‌های نژادی/قومی در شیوع مصرف سیگار در سطح ایالت‌ها پرداخت. مهم‌ترین یافته این تحقیق آن بود که **هیچ الگوی یکسانی برای این نابرابری‌ها در سراسر کشور وجود ندارد و این تفاوت‌ها به شدت وابسته به زمینه جغرافیایی هستند**. به عنوان مثال، در حالی که احتمال سیگار کشیدن در میان بزرگسالان هیستپانی در اکثر ایالت‌ها کمتر از سفیدپوستان بود، این الگو برای بزرگسالان سیاه‌پوست بسیار متغیر بود؛ در برخی ایالت‌ها شانس سیگار کشیدن آنها کمتر، در برخی بیشتر و در بسیاری بدون تفاوت معنادار بود. این تحقیق همچنین نشان داد که اگرچه شیوع کلی سیگار در اکثر ایالت‌ها کاهش یافته است، اما **شکاف بین گروه‌های نژادی/قومی در اکثر مناطق ثابت مانده است**.

همچنین، تحقیق برکوویتز و همکاران (Berkowitz et al., 2016) در ژورنال Cancer Epidemiology, Biomarkers & Prevention بر چالش "برآوردهای منطقه کوچک" (Small-Area Estimates - SAE) در داده‌های BRFSS (۲۰۱۲) تمرکز کرد. آنها با توسعه یک مدل رگرسیون لجستیک چندسطحی (Multilevel Logistic Regression Model)، توانستند برآوردهای دقیقی از شش رده وضعیت سیگار کشیدن را در **سطح شهرستان** برای کل ایالات متحده تولید کنند. یافته‌های آنها **واریانس و تفاوت‌های بسیار زیادی** را در شیوع مصرف و ترک سیگار، نه تنها بین ایالت‌ها، بلکه در بین شهرستان‌های درون یک

ایالت، آشکار ساخت (مانند تمرکز بالاترین شیوع سیگار در ایالت‌های منطقه آپالاشیا). آنها نتیجه گرفتند که برای برنامه‌ریزی مؤثر مداخلات کنترل دخانیات، تکیه بر آمارهای ملی یا ایالتی کافی نیست و نیاز به داده‌های دقیق و محلی در سطح شهرستان داریم.

جایگاه تحقیق حاضر در مقایسه با مطالعات فوق:

تحقیق ما، ضمن تأیید اهمیت نابرابری‌های جمعیتی و واریانس جغرافیایی که توسط میلز و برکوویتز برجسته شده، یک گام فراتر می‌رود. در حالی که برکوویتز بر تولید برآوردهای دقیق جغرافیایی (**descriptive estimation**) تمرکز دارند، هدف ما ساخت "مدل‌های پیش‌بینی‌کننده" است. یافته ما مبنی بر اهمیت ویژگی x_state در مدل‌های RandomForest، کاملاً با نتیجه‌گیری آنها در مورد اهمیت واریانس جغرافیایی همسو است. مدل ما به صورت غیرمستقیم یاد گرفته است که "ایالت" به عنوان یک نماینده برای مجموعه‌ای از عوامل سیاستی، فرهنگی و اقتصادی عمل می‌کند که بر ترک سیگار مؤثرند. تحقیق آنها به ما می‌گوید که شیوع بالاتر یا پایین‌تر است، اما ما با استفاده از مدل‌های بوستینگ گرادیان (مانند LightGBM) (به طور ذاتی قادر به یادگیری تعاملات پیچیده بین ویژگی‌های فردی و زمینه‌ای -مانند ایالت- هستند) و تکنیک SHAP، تلاش می‌کنیم تا توضیح دهیم که چرا این اتفاق می‌افتد، یعنی کدام ویژگی‌های فردی و سلامتی در تعامل با این زمینه جغرافیایی، منجر به ترک یا ادامه مصرف می‌شوند.

5.3. عوامل مرتبط با سلامت و فرضیه شوک سلامتی

نقش وضعیت سلامت به عنوان یک عامل تعیین‌کننده در ترک سیگار، به طور گسترده در ادبیات علمی مورد بررسی قرار گرفته است. "فرضیه شوک سلامتی" بیان می‌کند که تشخیص یک بیماری جدی می‌تواند به عنوان یک کاتالیزور قدرتمند برای تغییرات رفتاری مثبت عمل کند. تحقیقات متعددی نشان داده‌اند که تشخیص بیماری‌های مزمن، به‌ویژه بیماری‌های مستقیماً مرتبط با سیگار، یکی از قوی‌ترین پیش‌بینی‌کننده‌های ترک موفق است.

مطالعه ای برجسته و مرتبط در این زمینه، تحقیق پاریک و همکاران (Parikh et al., 2022) است که در ژورنال *Stroke* منتشر شده است. این پژوهش با تحلیل داده‌های تجمیع‌شده BRFSS (2013-2019)، به بررسی جامع اپیدمیولوژی ترک سیگار در میان بازماندگان سکته مغزی پرداخته است. هدف اصلی آن‌ها، مقایسه نرخ ترک در این گروه با بازماندگان سرطان بود، با این فرض که بازماندگان سرطان به دلیل وجود طرح‌های ملی هدفمند، مداخلات ترک سیگار مؤثرتری دریافت می‌کنند. یافته‌های اصلی آنها نشان داد که با وجود اینکه سکته مغزی یک "شوک سلامتی" بزرگ است، اما بازماندگان آن به طور معناداری کمتر از بازماندگان سرطان سیگار را ترک می‌کنند (نسبت شانس ترک: ۰.۷۲؛ ۹۵٪ CI, 0.67–0.79)، حتی پس از کنترل سایر عوامل دموگرافیک و بیماری‌های همراه.

علاوه بر این، این تحقیق نابرابری‌های دموگرافیک و جغرافیایی قابل توجهی را در نرخ ترک در میان بازماندگان سکته مغزی آشکار ساخت. به طوری که نرخ ترک در افراد جوان‌تر (>۶۰ سال) (۴۳.۳٪ در مقابل ۷۴.۶٪)، زنان (۵۷.۸٪ در مقابل ۶۳.۴٪)، و

ساکنان مناطق روستایی و ایالت‌های "کمربند سخته" پایین‌تر بود. این یافته‌ها بر نیاز فوری به مداخلات هدفمند برای این گروه پرخطر تأکید می‌کند.

جایگاه تحقیق حاضر در مقایسه با مطالعه پاریخ و همکاران:

تحلیل ما، ضمن تأیید یافته‌های کلیدی در مورد نقش بیماری‌ها، از چند جنبه مهم کار پاریخ و همکاران را تکمیل و تعمیق می‌بخشد:

1. **دامنه تحلیل:** در حالی که مطالعه آنها بر جمعیت خاصی از بازماندگان سکته مغزی و سرطان تمرکز دارد، تحقیق حاضر با بررسی **کل جمعیت سیگاری‌های سابق و فعلی**، یک مدل **عمومی‌تر و قابل تعمیم‌تر** از عوامل ترک سیگار ارائه می‌دهد که می‌تواند برای کل جمعیت بزرگسال کاربرد داشته باشد.

2. **نوآوری روش‌شناختی:** تحقیق آنها عمدتاً بر رگرسیون لجستیک استوار است. در مقابل، ما با به‌کارگیری مجموعه‌ای از مدل‌های یادگیری ماشین غیرخطی (مانند LightGBM) و تکنیک تفسیرپذیری SHAP، توانسته‌ایم روابط پیچیده و تعاملی را که مدل‌های خطی قادر به کشف آن‌ها نیستند، آشکار سازیم. برای مثال، در حالی که یافته‌های آنها نشان می‌دهد که سن بالاتر با ترک بیشتر مرتبط است (یک رابطه خطی)، تحلیل ما با SHAP نشان داد که این رابطه پیچیده‌تر بوده و در سنین بسیار بالا، به یک **مانع** تبدیل می‌شود.

3. **پایداری زمانی و به‌روزرسانی:** تحقیق ما با تحلیل داده‌ها تا سال ۲۰۲۳، یک دیدگاه به‌روزتر و طولی از پایداری این الگوها، به خصوص در دوره پرتلاطم پس از سال ۲۰۱۹، ارائه می‌دهد و نشان می‌دهد که عوامل اصلی پیش‌بینی‌کننده با وجود تغییرات اجتماعی، به طور شگفت‌آوری ثابت باقی مانده‌اند.

بنابراین، ضمن تأیید یافته‌های اصلی در مورد اهمیت شوک سلامتی و وجود نابرابری‌ها، کار ما با استفاده از روش‌شناسی پیشرفته‌تر و دامنه وسیع‌تر، به درک دقیق‌تر و ظریف‌تری از دینامیک پیچیده حاکم بر ترک سیگار کمک می‌کند و بر ضرورت فراتر رفتن از تحلیل‌های جمعیتی خاص به سمت مدل‌های پیش‌بینی‌کننده عمومی تأکید دارد.

5.4. تحلیل روابط متقابل و نتایج ناخواسته در بهداشت عمومی

علاوه بر شناسایی پیش‌بینی‌کننده‌های مستقیم ترک سیگار، درک روابط پیچیده و گاهی متقابل بین رفتارهای مرتبط با سلامت، برای سیاست‌گذاری مؤثر ضروری است. یکی از مهم‌ترین این روابط، ارتباط بین سیگار کشیدن، ترک آن، و وزن بدن (BMI) است. با توجه به روند همزمان کاهش نرخ سیگار کشیدن و افزایش شیوع چاقی در بسیاری از کشورها، این سوال مطرح می‌شود که آیا سیاست‌های موفق کنترل دخانیات، به صورت ناخواسته به افزایش وزن در سطح جمعیت کمک کرده‌اند یا خیر.

یک مطالعه برجسته در این زمینه، تحقیق کاستریدیس و ین (Yen & Kasteridis, 2012) است که با استفاده از داده‌های BRFSS، به بررسی دقیق تأثیر ترک سیگار بر شاخص توده بدنی (BMI) پرداخته‌اند. هدف اصلی آنها، کمی‌سازی میزان

افزایش وزن پس از ترک و بررسی این بود که آیا این اثر بر اساس سن، جنسیت، و مدت زمان ترک متفاوت است یا خیر. آنها با استفاده از مدل‌های پیشرفته "اثر درمانی" (treatment effect models) و با به‌کارگیری مالیات ایالتی بر سیگار به عنوان یک متغیر ابزاری، توانستند "اثر خالص" ترک سیگار را با کنترل عوامل مخدوش‌کننده پنهان (endogeneity) تخمین بزنند. یافته‌های اصلی آنها نشان داد:

1. **رابطه مثبت و پایدار:** ترک سیگار به طور معناداری با **افزایش BMI** مرتبط است. این افزایش بلافاصله پس از ترک شروع شده و با گذشت زمان، به خصوص در ۵ تا ۱۰ سال اول، به اوج خود می‌رسد.

2. **تفاوت بین گروه‌ها:** این اثر در زنان (به‌ویژه زنان میانسال) و **افراد مسن‌تر** به مراتب قوی‌تر از جوانان بود.

جایگاه تحقیق حاضر در مقایسه با مطالعه کاستریدیس و ین:

تحقیق کاستریدیس و ین یک چارچوب تحلیلی مهم برای تفسیر یکی از ویژگی‌های کلیدی در مدل ما، یعنی BMI، فراهم می‌کند. در حالی که کار آنها بر BMI به عنوان **متغیر وابسته (outcome)** تمرکز دارد، تحقیق ما BMI را به عنوان یک **متغیر مستقل (predictor)** برای ترک سیگار در نظر می‌گیرد. این دو دیدگاه، دو روی یک سکه هستند و یافته‌های ما به نوعی مکمل کار آنهاست:

- در تحلیل SHAP ما، متغیر x_{bmi5} به طور مداوم به عنوان یکی از ۱۰ ویژگی برتر ظاهر می‌شود، اما تأثیر آن پیچیده و **غیرخطی** است. ابر نقاط در نمودار SHAP نشان می‌دهد که هم مقادیر بسیار پایین BMI و هم مقادیر بسیار بالا، می‌توانند به عنوان مانعی برای ترک عمل کنند.
- این یافته می‌تواند با نتایج کاستریدیس و ین مرتبط باشد. ممکن است **نگرانی از افزایش وزن پس از ترک** (که آنها میزان آن را کمی‌سازی کردند)، برای افرادی که از قبل با وزن خود مشکل دارند (چه کمبود وزن و چه اضافه‌وزن)، به یک **مانع روانی مهم** برای اقدام به ترک تبدیل شود.

5.5. دینامیک زمانی و تأثیر رویدادهای کلان: پاندمی کووید-۱۹

یکی از پرسش‌های کلیدی در تحلیل طولی داده‌های بهداشت عمومی، بررسی تأثیر رویدادهای بزرگ اجتماعی بر الگوهای رفتاری است. پاندمی کووید-۱۹، با ایجاد استرس‌های روانی، اقتصادی و اختلال در سیستم‌های درمانی، یک زمینه منحصر به فرد برای مطالعه پایداری یا تغییر در رفتارهای مرتبط با ترک سیگار فراهم آورد. در حالی که انتظار می‌رفت افزایش آگاهی از خطرات تنفسی، افراد را به سمت ترک سوق دهد، شواهد نشان‌دهنده یک تصویر پیچیده‌تر است.

- مطالعه‌ای برجسته و مرتبط در این زمینه، تحقیق **باندی و همکاران (Bandi et al., 2022)** است که در ژورنال JAMA Network Open منتشر شده است. این پژوهش به طور خاص به بررسی **تغییرات در رفتارهای مرتبط با ترک سیگار در طول پاندمی کووید-۱۹** پرداخته است. نوآوری کلیدی این تحقیق، استفاده از دو منبع داده مستقل و مکمل

(داده‌های نظرسنجی BRFSS و داده‌های خرده‌فروشی NielsenIQ برای ردیابی فروش محصولات جایگزین نیکوتین) برای اعتبارسنجی یافته‌ها بود. یافته‌های اصلی آنها یک سیگنال واضح و هشداردهنده را نشان داد: برای اولین بار از سال ۲۰۱۱، شیوع "تلاش برای ترک" در میان سیگاری‌ها بین سال‌های ۲۰۱۹ و ۲۰۲۰ به طور معناداری کاهش یافت (از ۶۵.۲٪ به ۶۳.۲٪). این کاهش به‌ویژه در گروه‌هایی که بیشترین آسیب را از پاندمی دیدند (مانند افراد با بیماری‌های همراه، زنان و اقلیت‌های نژادی) برجسته‌تر بود. همزمان، فروش محصولات NRT در مقایسه با روند مورد انتظار، ۱٪ تا ۱۳٪ کاهش داشت که این کاهش فعالیت در زمینه ترک را تأیید می‌کند. نویسندگان نتیجه گرفتند که فعالیت جدی برای ترک سیگار در میان بزرگسالان آمریکایی بلافاصله پس از شروع پاندمی کاهش یافت و این امر بر نیاز فوری برای درگیر کردن مجدد سیگاری‌ها با استراتژی‌های ترک مبتنی بر شواهد تأکید می‌کند.

جایگاه تحقیق حاضر در مقایسه با مطالعه باندی و همکاران:

تحلیل ما، که داده‌های آن تا سال ۲۰۲۳ را پوشش می‌دهد، می‌تواند یافته‌های باندی و همکاران را تأیید، تکمیل و تعمیق بخشد:

- **تأیید پایداری با وجود نوسانات:** در حالی که تحقیق آنها یک کاهش کوتاه‌مدت در "تلاش برای ترک" در سال ۲۰۲۰ را نشان می‌دهد، یافته کلیدی پروژه ما این است که الگوهای اصلی "موفقیت در ترک" در طول زمان (۲۰۱۸-۲۰۲۳) به طور شگفت‌آوری ثابت مانده‌اند. این دو یافته متناقض نیستند؛ بلکه یک داستان کامل‌تر را روایت می‌کنند: ممکن است پاندمی به طور موقت تعدد تلاش‌ها را کاهش داده باشد، اما عوامل بنیادینی که تعیین می‌کنند یک تلاش موفق خواهد بود یا خیر (مانند سن، تاهل، شوک سلامتی)، ثابت باقی مانده‌اند.
 - **نوآوری روش‌شناختی:** ما با استفاده از مدل‌های یادگیری ماشین پیشرفته، نه تنها "آیا" الگوها تغییر کرده‌اند را بررسی می‌کنیم، بلکه "چگونه" و "کدام" عوامل بیشترین تأثیر را داشته‌اند. تحلیل SHAP ما نشان می‌دهد که حتی در سال‌های پاندمی، سلسله مراتب اهمیت ویژگی‌ها تغییر چشمگیری نکرده است.
 - **مدل جامع به عنوان راه حل:** یافته‌های باندی و همکاران بر ضرورت "درگیر کردن مجدد" سیگاری‌ها تأکید می‌کند. مدل پیش‌بینی‌کننده جامعی که ما بر اساس داده‌های تجمیع‌شده پیشنهاد می‌دهیم، می‌تواند به عنوان ابزاری برای شناسایی گروه‌های پرخطر و شخصی‌سازی این تلاش‌ها برای درگیر کردن مجدد، مورد استفاده قرار گیرد.
- بنابراین، تحقیق ما با ارائه یک دیدگاه بلندمدت‌تر و با استفاده از روش‌شناسی پیشرفته‌تر، نشان می‌دهد که با وجود نوسانات رفتاری کوتاه‌مدت ناشی از رویدادهای بزرگ مانند پاندمی، عوامل اصلی و ساختاری مؤثر بر موفقیت در ترک سیگار، ریشه‌دار و پایدار هستند. این بر اهمیت تمرکز بر این عوامل بنیادین در سیاست‌گذاری‌های بلندمدت بهداشت عمومی تأکید می‌کند.

6. بحث (Discussion)

این پژوهش با هدف توسعه و ارزیابی مدل‌های یادگیری ماشین برای پیش‌بینی ترک موفق سیگار در میان بزرگسالان با استفاده از داده‌های گسترده پیمایش رفتاری و سلامت (BRFSS) در یک دوره شش ساله (۲۰۱۸ تا ۲۰۲۳) انجام شد. فراتر از صرفاً ساخت مدل‌های پیش‌بینی، هدف اصلی، دستیابی به یک درک عمیق‌تر، پایدارتر و قابل‌تفسیر از عوامل تأثیرگذار بر این پدیده بهداشت عمومی حیاتی بود. یافته‌های ما الگوهای پیچیده‌ای از ارتباط میان مشخصه‌های فردی، رفتاری و جغرافیایی را با موفقیت در ترک سیگار آشکار می‌سازند که بسیاری از آن‌ها در طول زمان، پایداری چشمگیری نشان داده‌اند.

6.1. خلاصه و تفسیر یافته‌های اصلی

یافته‌های این مطالعه نشان می‌دهد که با بهره‌گیری از مدل‌های یادگیری ماشین، می‌توان به دقت قابل قبولی در پیش‌بینی ترک سیگار دست یافت و عوامل مؤثر بر آن را شناسایی کرد:

- **برتری مدل‌های غیرخطی در پیش‌بینی:** مدل‌های مبتنی بر گرادیان بوستینگ، به ویژه LightGBM و CatBoost، به طور مداوم بهترین عملکرد پیش‌بینی را از نظر AUC و F1-Score در تمامی سال‌ها از خود نشان دادند. این امر بر پیچیدگی روابط میان عوامل متعدد و ترک سیگار تأکید دارد، که مدل‌های خطی سنتی مانند رگرسیون لجستیک و LDA (علی‌رغم دقت کلی مناسب در برخی سال‌ها) قادر به درک کامل آن نیستند. برتری مداوم سناریوی "تمام ویژگی‌های گلچین شده" نسبت به ویژگی‌های استخراج شده توسط روش VIF نیز این ایده را تقویت می‌کند که حتی با حضور هم‌خطی، مجموعه‌ی کامل ویژگی‌های گلچین شده، اطلاعات ارزشمندتری را برای مدل‌های پیچیده یادگیری ماشین فراهم می‌آورند.
- **قدرت پیش‌بینی و رویکردهای جامع (Pooled Models & Ensemble):** برتری مدل‌های مبتنی بر گرادیان بوستینگ (CatBoost و LightGBM) در تمامی تحلیل‌ها (اعم از سالانه یا جامع) قویاً بر پیچیدگی روابط میان عوامل متعدد و ترک سیگار تأکید دارد، که مدل‌های خطی قادر به درک کامل آن نیستند. بالاترین عملکرد (به ویژه از نظر AUC)، توسط رویکردهای مدل‌سازی جامع و ترکیبی یعنی **تجمیع داده (Data Pooling)**، **آنسامبل زمانی (Temporal Ensemble)** و روش انباشت (**Stacking Ensemble**) به دست آمد، با **Stacking Ensemble** که به طور خاص **بهترین عملکرد کلی (AUC: 0.7881) را ثبت کرد**. این یافته‌ها بر مزایای تجمیع اطلاعات در طول زمان و استفاده از توان ترکیبی مدل‌های پیشرفته برای پیش‌بینی‌های پایدارتر و دقیق‌تر تأکید می‌کند.
- **ثبات و تأثیرات پایدار عوامل پیش‌بین:** تحلیل SHAP و ضرایب رگرسیون لجستیک، مجموعه‌ای از پیش‌بین‌های کلیدی را با تأثیر ثابت در طول دوره مطالعه (۲۰۱۸-۲۰۲۳) شناسایی کرد:
 - **سن:** مهم‌ترین و پایدارترین پیش‌بین بود، به طوری که **افزایش سن، به طور مداوم و قاطعانه با افزایش احتمال ترک موفق سیگار همراه بود**.
 - **وضعیت تأهل:** متأهل بودن به طور پایدار با افزایش شانس ترک سیگار مرتبط بود، در حالی که سایر وضعیت‌های تأهل (ازدواج نکرده، طلاق‌گرفته، جدا شده یا بیوه) احتمال ترک سیگار را کاهش می‌داد.

- **وضعیت اقتصادی-اجتماعی:** سطوح بالاتر درآمد و تحصیلات به طور مداوم و مثبت با احتمال ترک موفق سیگار همراه بوده‌اند.
- **ویژگی‌های فیزیکی (BMI) و وزن:** هر دو BMI بالاتر و وزن بیشتر به طور پایدار با افزایش احتمال ترک موفق سیگار مرتبط بودند.

• سلامت و بیماری‌ها: هم تحریک‌کننده و هم بازدارنده:

- **تشخیص بیماری‌های جدی (مانند COPD، سکته مغزی، یا تشخیص افسردگی):** همانطور که در فرضیات اولیه پژوهش (**فرضیه شوک سلامتی**) انتظار می‌رفت، این شرایط پزشکی جدی به طور قاطع با افزایش قابل توجه احتمال ترک موفق سیگار مرتبط بودند. این نشان می‌دهد که دریافت یک تشخیص پزشکی نگران‌کننده می‌تواند به عنوان یک شوک یا انگیزه‌ی قوی برای ترک سیگار عمل کند.
- **سلامت عمومی:** افرادی که وضعیت سلامت عمومی **بدتری** را گزارش می‌کردند، به طور پایدار، احتمال کمتری برای ترک موفق سیگار داشتند. این تفاوت در نتایج (بین "سلامت عمومی بدتر" که مانع است، و "تشخیص بیماری جدی" که محرک است) نشان می‌دهد که اگرچه یک بیماری شدید خاص می‌تواند انگیزه ایجاد کند، اما در کل، کیفیت پایین عمومی زندگی و سلامتی ممکن است توانایی فرد را برای انجام تغییرات مثبت در سبک زندگی (مانند ترک سیگار) تضعیف کند.
- **سلامت روان نامناسب:** همانطور که انتظار می‌رفت، **مشکلات سلامت روان (تعداد روزهای بیشتر با سلامت روان نامناسب)** به طور پیوسته، و به ویژه از سال ۲۰۱۹/۲۰۲۰ به بعد که الگو تثبیت شد، با کاهش احتمال ترک موفق سیگار مرتبط بود. این بر نقش حیاتی سلامت روان در روند ترک سیگار تأکید می‌کند و نشان می‌دهد که افسردگی یا سایر مشکلات روانی می‌توانند به عنوان یک **مانع قابل توجه** عمل کنند.

• سبک زندگی و چکاپ پزشکی:

- **فعالیت فیزیکی:** افرادی که فعالیت فیزیکی و ورزش انجام می‌دهند به طور پیوسته احتمال ترک موفق سیگار بیشتری داشتند، در حالی که افراد غیر فعال شانس کمتری داشتند.
- **چکاپ پزشکی:** انجام منظم چکاپ پزشکی (فاصله زمانی کمتر تا آخرین چکاپ) به طور مداوم با افزایش شانس ترک موفق سیگار همراه بود که بر نقش پیشگیرانه و مشاوره پزشکی در این فرآیند تأکید دارد.

6.2. مقایسه با تحقیقات پیشین

یافته‌های این مطالعه با استفاده از تحلیل جامع داده‌های (BRFSS ۲۰۱۸-۲۰۲۳) و مدل‌های پیشرفته یادگیری ماشین، ضمن تأیید برخی مفاهیم شناخته‌شده در ادبیات کنترل دخانیات، بینش‌های ظریف‌تر و پویاتری را ارائه می‌دهد

6.2.1. عوامل محیطی و سیاست‌گذاری در سطح ایالت

تحقیقات قبلی بر نقش حیاتی محیط سیاست، به ویژه سیاست‌های مالیاتی بر سیگار، در تغییر رفتارهای سیگار کشیدن تأکید کرده‌اند. مطالعاتی مانند ژو و همکاران (Zhu et al., 2025) و شاربا و همکاران (Sharbaugh et al., 2018) به طور مستقیم افزایش مالیات بر سیگار را با افزایش نرخ ترک سیگار و کاهش شیوع کلی آن مرتبط دانسته‌اند. یافته‌های این پژوهش، اگرچه مستقیماً اثر سیاست مالیاتی را اندازه‌گیری نمی‌کنند، اما اهمیت بالایی متغیر x_state (ایالت) را در تحلیل‌های RandomForest Feature Importance و نمودارهای SHAP به طور مداوم نشان دادند. این نشان می‌دهد که مدل‌های پیشرفته ما، به صورت غیرمستقیم، در حال یادگیری تأثیرات ناشی از تفاوت‌های سیاستی، قانونی، فرهنگی، و خدماتی میان ایالت‌های مختلف هستند که در داده‌های سطح ایالت کدگذاری شده و بر ترک سیگار تأثیر می‌گذارند. در حالی که مطالعات ژو و شاربا بر تأثیر سیاست‌ها بر «شیوع» و «تلاش برای ترک» تمرکز دارند، مطالعه حاضر بر «ترک موفق سیگار» در میان کسانی که سابقه سیگار کشیدن دارند، تمرکز می‌کند. این امر دیدگاهی جامع‌تر از فرآیند تغییر رفتار در افراد سیگاری و ارتباط آن با زمینه ایالتی فراهم می‌آورد.

6.2.2. عوامل محیطی و نابرابری‌های جمعیتی در سطح ایالت

مطالعات پیشین بر نابرابری‌های جمعیتی و واریانس جغرافیایی در شیوع سیگار تأکید کرده‌اند (Mills et al., 2021)؛ اهمیت پایدار متغیر x_state در تحلیل RandomForest Feature Importance و نمودارهای SHAP ما، قویاً این نتیجه‌گیری را در مورد اهمیت تفاوت‌های جغرافیایی تأیید می‌کند. این نشان می‌دهد که «ایالت» به عنوان یک متغیر تجمعی، گستره‌ای از عوامل سیاستی، اقتصادی، و فرهنگی مؤثر بر ترک سیگار را نمایندگی می‌کند. پژوهش ما، با استفاده از مدل‌های بوستینگ و تحلیل SHAP، فراتر از تولید برآوردهای توصیفی می‌رود و به این پرسش می‌پردازد که چگونه ویژگی‌های فردی در تعامل با این زمینه جغرافیایی بر تصمیم‌گیری ترک تأثیر می‌گذارند.

6.2.3. عوامل مرتبط با سلامت و فرضیه شوک سلامتی

نقش وضعیت سلامت به عنوان یک عامل تعیین‌کننده در ترک سیگار به طور گسترده مورد بررسی قرار گرفته است. مطالعه پاریک و همکاران (Parikh et al., 2022) بر اهمیت این شوک در بازماندگان سکته مغزی تأکید می‌کند. تحلیل ما، ضمن تأیید یافته‌های کلیدی در مورد نقش بیماری‌ها، دامنه این بررسی را گسترش داده و آن را در جمعیت کلی سیگاری‌های بزرگسال مورد بررسی قرار می‌دهد. یافته‌های ما به وضوح نشان دادند که تشخیص بیماری‌های مزمن جدی مانند COPD، سابقه سکته مغزی، و تشخیص اختلال افسردگی، به طور قاطع و پایدار با افزایش قابل توجه احتمال ترک موفق سیگار مرتبط بودند. این نتایج قویاً «فرضیه شوک سلامتی» را تأیید می‌کند. در مقایسه با مطالعه پاریک و همکاران که بر جمعیت خاصی (بازماندگان سکته مغزی و سرطان) و عمدتاً بر رگرسیون لجستیک متمرکز است، تحقیق ما یک مدل پیش‌بینی‌کننده عمومی‌تر و قابل تعمیم‌تر را با استفاده از مدل‌های یادگیری ماشین غیرخطی (LightGBM) و تکنیک تفسیرپذیری SHAP ارائه می‌دهد. این رویکرد به ما امکان داد تا روابط پیچیده‌تر و تعاملی را که مدل‌های خطی ممکن است نادیده بگیرند، کشف کنیم.

6.2.4. تحلیل روابط متقابل: سیگار کشیدن، ترک و وزن بدن و BMI

درک روابط پیچیده و گاهی غیرمستقیم میان رفتارهای مرتبط با سلامت، برای سیاست‌گذاری مؤثر حیاتی است. ارتباط میان ترک سیگار و تغییر وزن (BMI) یکی از این جنبه‌هاست. مطالعه کاستریدیس و ین (Yen, 2012 & Kasteridis) به بررسی افزایش BMI پس از ترک سیگار پرداخت و نشان داد که این افزایش می‌تواند پایدار باشد و در برخی گروه‌ها شدیدتر است. تحقیق حاضر، BMI را به عنوان یک «پیش‌بین» ترک سیگار مورد بررسی قرار داد، در حالی که مطالعه کاستریدیس و ین آن را به عنوان یک «پیامد» ترک در نظر گرفت. این دو دیدگاه مکمل یکدیگر هستند: تحلیل SHAP ما نشان داد که BMI بالاتر و وزن بیشتر، به طور پیوسته با افزایش احتمال ترک موفق سیگار مرتبط هستند. این یافته با کار کاستریدیس و ین متناقض نیست؛ در عوض، ممکن است نشان دهد که افراد دارای BMI بالاتر (و احتمالاً با عوارض جانبی سلامتی مرتبط) ممکن است انگیزه‌های قوی‌تری برای ترک سیگار داشته باشند (به عنوان مثال، از طریق تشدید شوک سلامتی یا توصیه‌های پزشکی). در حالی که مطالعه کاستریدیس و ین نشان می‌دهد ترک سیگار می‌تواند به افزایش وزن منجر شود، یافته‌های ما یک عامل بالقوه (BMI بالا) را شناسایی می‌کند که ممکن است فرد را به سمت اقدام به ترک سوق دهد. این ارتباط چندوجهی، پیچیدگی متقابل رفتارهای سلامت و نیاز به رویکردهای جامع برای مدیریت سلامت را نشان می‌دهد.

6.2.5. دینامیک زمانی و تأثیر رویدادهای کلان: پاندمی کووید-۱۹

برخی تحقیقات، از جمله مطالعه باندی و همکاران (Bandi et al., 2022)، کاهش کوتاه‌مدت در «تلاش برای ترک سیگار» در طول پاندمی کووید-۱۹ (بین سال‌های ۲۰۱۹ و ۲۰۲۰) را نشان دادند. در مقابل، یافته کلیدی و مهم مطالعه حاضر، برتری مدل‌های بوستینگ و پایداری عوامل اصلی پیش‌بینی‌کننده موفقیت در ترک سیگار (نظیر سن، وضعیت تأهل، و شوک سلامتی) در طول دوره جامع ۲۰۱۸ تا ۲۰۲۳، با وجود نوسانات اجتماعی بزرگ ناشی از پاندمی، است. این نشان می‌دهد که اگرچه رویدادهای کلان می‌توانند به طور موقت بر تمایل افراد برای «شروع تلاش» جهت ترک سیگار تأثیر بگذارند، اما عوامل بنیادینی که تعیین می‌کنند یک تلاش موفقیت‌آمیز خواهد بود یا خیر، ریشه‌دار و پایدار باقی می‌مانند.

6.3. نقاط قوت و محدودیت‌ها

نقاط قوت:

- **داده‌های ملی و حجیم:** استفاده از داده‌های BRFSS، یک منبع معتبر و ملی، قابلیت تعمیم‌پذیری یافته‌ها به جمعیت بزرگسال ایالات متحده را افزایش می‌دهد.
- **پوشش زمانی جامع:** تحلیل داده‌ها در طول یک دوره شش ساله (۲۰۱۸-۲۰۲۳)، ارزیابی پایداری عوامل پیش‌بینی‌کننده را امکان‌پذیر ساخته و بینش‌های مهمی در مورد روندهای ثابت و متغیر ارائه می‌کند. این رویکرد طولی بر روی داده‌های مقطعی تکرارشونده، به تشخیص الگوهای مقاوم در برابر نوسانات محیطی کمک کرده است.

- استفاده از روش‌های پیشرفته یادگیری ماشین: بهره‌گیری از مدل‌های بوستینگ به‌ویژه CatBoost و LightGBM و ابزارهای تفسیرپذیری مانند SHAP، این مطالعه را قادر ساخت تا روابط غیرخطی پیچیده‌تر و تعاملات پنهان میان ویژگی‌ها را شناسایی و تأثیر آن‌ها را شفاف‌سازی کند که فراتر از توان مدل‌های آماری خطی است.
- رویکردهای مدل‌سازی جامع: استفاده از رویکردهای پیشرفته مانند تجميع داده، آنسامبل زمانی و به ویژه Stacking Ensemble که توانایی یادگیری الگوها از مجموعه‌ی کامل داده‌های طولی را فراهم آورده و منجر به بهبود چشمگیر دقت و قابلیت تعمیم‌پذیری مدل‌های پیش‌بینی‌کننده شد. این رویکرد به ویژه در مدیریت داده‌های حجیم و متغیر در طول زمان کارایی خود را ثابت کرد.

محدودیت‌ها:

- ماهیت خودگزارش‌دهی: داده‌های BRFSS مبتنی بر گزارش شخصی پاسخ‌دهندگان است که می‌تواند منجر به سوگیری یادآوری (Recall Bias) و یا سوگیری تمایل اجتماعی (Social Desirability Bias) شود (مثلاً تمایل به گزارش ترک سیگار حتی اگر واقعاً چنین نبوده است).
- ماهیت متغیر هدف و تعاریف: تعریف "ترک موفق سیگار" بر اساس "سیگاری سابق"، گرچه رایج است، اما ممکن است طیف کامل وضعیت‌های ترک را شامل نشود (مثلاً شامل "ترک‌های موقت" نیست) یا برای مطالعه تماماً طولی افراد (از یک موج به موج بعدی) مستقیماً مناسب نبوده است.
- عدم توازن کلاس‌ها: شیوع بالاتر کلاس "ترک موفق سیگار" (اکثریت) در نمونه مطالعه، منجر به سوگیری مدل‌ها به سمت پیش‌بینی کلاس اکثریت شد، به طوری که مدل PCA+Logistic Regression عملاً در تشخیص کلاس "ادامه مصرف سیگار" ناموفق بود و حتی مدل‌های بوستینگ نیز True Positives بسیار بالاتری نسبت به True Negatives داشتند. این مسئله، هرچند که در حوزه پیش‌بینی "ترک موفق" قدرت مدل را نشان می‌دهد، اما می‌تواند کاربرد آن را برای شناسایی افراد در "خطر عدم موفقیت در ترک" محدود کند.
- محدودیت‌های متغیرهای موجود: اگرچه BRFSS جامع است، اما ممکن است تمام عوامل رفتاری، روانی-اجتماعی یا محیطی با جزئیات کافی برای یک مدل‌سازی جامع و تفسیری در مورد ترک سیگار نباشد (مانند سطوح دقیق نیکوتین، حمایت‌های اجتماعی غیر از وضعیت تأهل، یا جزئیات برنامه‌های ترک سیگار که فرد در آن‌ها شرکت کرده است).

6.4. پیامدها و مسیرهای آتی پژوهش

- پیامدهای عملیاتی و بهداشتی: یافته‌های ما برای طراحی مداخلات هدفمندتر و شخصی‌سازی شده برای ترک سیگار پیامدهای مهمی دارند:

- **هدف گذاری (Targeting) بر اساس سن:** لزوم بازنگری در برنامه‌های ترک سیگار که سن بالاتر را به عنوان یک مانع در نظر می‌گیرند. در عوض، سن بالا می‌تواند یک پنجره فرصت (فراتر از تنها سلامت، از نظر پختگی تصمیم‌گیری یا درک تبعات طولانی مدت) باشد که نیاز به رویکردهای حمایتی مناسب با این گروه سنی دارد.
- **اهمیت سلامت روان:** نتایج بر ضرورت رویکرد یکپارچه سلامت جسم و روان در برنامه‌های ترک سیگار تأکید می‌کنند. ارزیابی و پشتیبانی فعال از سلامت روان (به ویژه در افراد دارای مشکلات روانی) می‌تواند شانس ترک سیگار را به طور قابل ملاحظه‌ای افزایش دهد.
- **نقش آگاهی پزشکی و چکاپ‌های منظم:** تقویت دسترسی و تشویق به چکاپ‌های منظم پزشکی و مشاوره سلامت می‌تواند به عنوان یک ابزار حیاتی برای افزایش شانس ترک سیگار (و احتمالاً از طریق ایجاد نقاط بحرانی آگاهی‌بخش) عمل کند.
- **ملاحظات مرتبط با BMI و فعالیت:** نیاز به تحقیقات بیشتر برای درک پیچیدگی ارتباط BMI، وزن و فعالیت فیزیکی با ترک سیگار. این یافته‌های ظاهراً "متناقض" می‌توانند زمینه‌ای برای توسعه مداخلات چندوجهی باشند که سلامت فیزیکی را نیز در بر می‌گیرند.

• پیشنهادات برای تحقیقات آتی:

- **داده‌های طولی واقعی:** انجام مطالعات آینده با استفاده از داده‌های طولی واقعی که افراد را در طول زمان دنبال می‌کنند با امکان شناسایی ID تکرار شونده در سال‌ها، می‌تواند به شناسایی مسیرهای علت و معلولی کمک شایانی کند و فراتر از روابط همبستگی پیش‌بینی‌گرانه عمل کند.
- **مداخلات چند وجهی:** بررسی اثربخشی مداخلات جامع‌تر که هم‌زمان به عوامل دموگرافیک، سلامت جسم، سلامت روان و حتی مسائل مالی می‌پردازند، با توجه به اثرگذاری هم‌زمان این ویژگی‌ها در مدل‌های ما.
- **استفاده از تکنیک‌های مدیریت عدم توازن کلاس‌ها:** به‌کارگیری الگوریتم‌های پیشرفته‌تر مدیریت عدم توازن کلاس مانند SmoteBoost یا استفاده از Cost-sensitive Learning برای بهبود توانایی مدل‌ها در شناسایی افراد "ادامه دهنده سیگار" (کلاس اقلیت) می‌تواند قابلیت کاربردی مدل‌ها را به شدت ارتقاء بخشد.
- **کاوش عمیق‌تر عوامل جغرافیایی:** بررسی جزئی‌تر تأثیر عوامل جغرافیایی (مثلاً در سطح شهرستان یا محله) و تأثیر سیاست‌های محلی بر ترک سیگار می‌تواند به توسعه مداخلات مبتنی بر منطقه کمک کند. همچنین، آزمون صریح‌تر فرضیه‌ی اثر تعاملی بین سلامت روان و جغرافیا (که در این مطالعه به دلیل نبود خروجی صریح از این اصطلاحات در داده‌ها به طور کامل بررسی نشد) در پژوهش‌های آتی مفید خواهد بود.
- **بهره‌برداری از مدل‌های جامع:** ادامه تحقیق بر روی این مدل‌های ترکیبی برای کشف بینش‌های عمیق‌تر از تعاملات پیچیده عوامل مؤثر و پایداری زمانی آن‌ها در جمعیت‌های متنوع‌تر. این مدل‌ها به دلیل عملکرد برتر خود در پیش‌بینی، ابزار ایده‌آلی برای آزمایش فرضیه‌های جدید در محیط‌های پویا هستند.

- **تحلیل متامدل در Stacking:** انجام تحلیل‌های عمیق‌تر بر روی فرامدل (meta-model) در رویکرد Stacking برای درک بهتر چگونگی ترکیب دانش از مدل‌های پایه توسط این فرامدل و روشن ساختن روابط پیچیده‌تر.

7. نتیجه‌گیری (Conclusion)

این تحقیق با بهره‌گیری از مدل‌های یادگیری ماشین پیشرفته بر روی داده‌های (BRFSS ۲۰۱۸-۲۰۲۳)، با موفقیت توانسته است عوامل کلیدی و پایداری را در پیش‌بینی ترک موفق سیگار شناسایی کند. برتری مداوم مدل‌های بوستینگ غیرخطی نشان‌دهنده پیچیدگی روابطی است که رویکردهای سنتی از شناسایی آن‌ها بازمی‌مانند. یافته‌هایی چون تأثیر مثبت سن بالا، متأهل بودن و وضعیت اجتماعی-اقتصادی بالاتر، در کنار نقش دوگانه‌ی سلامت (محرک بودن تشخیص بیماری جدی و بازدارنده بودن سلامت عمومی پایین و سلامت روان نامناسب)، بر ضرورت سیاست‌گذاری‌ها و مداخلات شخصی‌سازی شده و جامع تأکید دارند. این نتایج داده‌محور، نه تنها دانش ما را در زمینه ترک سیگار غنی می‌سازند، بلکه ابزارهایی تحلیلی برای طراحی برنامه‌های مؤثرتر در بهداشت عمومی و درمان اعتیاد فراهم می‌آورند.

منابع و لینک‌ها (references)

- 1.1. Association of a major tobacco tax increase in California with increased smoking cessation (Shu-Hong Zhu et al., 2025)

Link: <https://academic.oup.com/jnci/advance-article/doi/10.1093/jnci/djaf121/8164449>

- 1.2. Impact of cigarette taxes on smoking prevalence from 2001-2015: A report using the Behavioral and Risk Factor Surveillance Survey (BRFSS) (Sharbaugh et al., 2018)

Link: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0204416>

- 2.1. State-Level Patterns and Trends in Cigarette Smoking Across Racial and Ethnic Groups in the United States, 2011–2018 (Mills et al., 2021)

Link: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8139454/>

- 2.2. Multilevel Small-Area Estimation of Multiple Cigarette Smoking Status Categories Using the 2012 Behavioral Risk Factor Surveillance System (Berkowitz et al., 2016)

Link: <https://aacrjournals.org/cebp/article/25/10/1402/70704/Multilevel-Small-Area-Estimation-of-Multiple>

3. **Smoking Cessation in Stroke Survivors in the United States: A Nationwide Analysis (Parikh et al., 2022)**

Link: <https://www.ahajournals.org/doi/full/10.1161/STROKEAHA.121.036941>

4. **Smoking Cessation and Body Weight: Evidence from the Behavioral Risk Factor Surveillance Survey (Kasteridis & Yen, 2012)**

Link: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3401400/>

5. **Changes in Smoking Cessation–Related Behaviors Among US Adults During the COVID-19 Pandemic**

Link: <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2794810>