فصلنامه علمی جامعه شناسی سیاسی انقلاب اسلامی
دوره ۵، شماره ٤، شماره پیاپی (۲۰)، زمستان ۱٤۰۳، ص ۱۸۸ ـ ۱۶۳
نوع مقاله: پژوهشی
شاپا الکترونیکی: ۵۲۰۰-۲۷۸۳
تاریخ دریافت: ۱۴۰۳/۱۰/۲۰
تاریخ پذیرش: ۱۴۰۳/۱۲/۲۸

بررسی عوامل مشارکت در انتخابات‌های جمهوری اسلامی ایران، با استفاده از الگوریتم‌های یادگیری عمیق و تکنیک‌های یادگیری گروهی

سجاد نیکخو[۱] میلاد فریدنیا[۲] کرم الله باقری فرد[۳]

چکیده:

مشارکت سیاسی به‌ویژه مشارکت در انتخابات یکی از مولفه‌های کلیدی در فرآیندهای دموکراتیک است که تحت تأثیر مجموعه‌ای از عوامل اجتماعی، اقتصادی و روان‌شناختی قرار دارد. در این تحقیق، به تحلیل و پیش‌بینی میزان مشارکت در انتخابات جمهوری اسلامی ایران با استفاده از مجموعه داده‌های حاصل از یک نظرسنجی عمومی پرداخته شده است. این مجموعه داده شامل ۱۰۰۰۰ رکورد از ویژگی‌های فردی و اجتماعی نظیر جنسیت، سن، سطح تحصیلات، وضعیت اقتصادی، شغلی و تعاملات رسانه‌ای می‌باشد. برای انجام تحلیل‌های پیش‌بینی، از مدل‌های پیشرفته یادگیری عمیق (شبکه‌های عصبی) و تکنیک‌های یادگیری گروهی مانند درختان تصادفی و تقویتی گرادیان بهره گرفته شده است. نتایج تجربی این تحقیق نشان می‌دهند که تحصیلات، درآمد ماهانه، باورهای سیاسی، تعاملات رسانه‌ای و سطح اعتماد به نظام سیاسی به‌طور معناداری بر تصمیم‌گیری افراد برای مشارکت در انتخابات تأثیر می‌گذارند. علاوه بر این، دقت مدل‌های یادگیری ماشین در پیش‌بینی مشارکت، با استفاده از ویژگی‌های اجتماعی و اقتصادی به میزان قابل توجهی افزایش یافته است. این تحقیق می‌تواند ابزار مفیدی برای تحلیلگران و طراحان سیاست‌های انتخاباتی باشد و به آن‌ها کمک کند تا استراتژی‌های موثرتری برای افزایش مشارکت و بهبود دقت پیش‌بینی‌های انتخاباتی اتخاذ کنند. همچنین، یافته‌های این مطالعه می‌تواند زمینه‌ساز مطالعات آینده در زمینه شبیه‌سازی و پیش‌بینی رفتارهای انتخاباتی باشد.

واژگان اصلی: مشارکت انتخاباتی، عوامل اجتماعی، یادگیری عمیق، یادگیری گروهی، تحلیل داده‌های اجتماعی.

۱. استادیار گروه معارف اسلامی دانشگاه فرهنگیان، تهران، ایران (نویسنده مسئول) nikkhoo313@yahoo.com
۲. دانشجوی دکتری، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی، یاسوج، ایران
۳. استادیار، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی، یاسوج، ایران

۱- مقدمه

مشارکت سیاسی، به‌ویژه شرکت در انتخابات، یکی از ارکان بنیادین هر نظام دموکراتیک است که نشان‌دهنده عمق مشروعیت و میزان تعامل میان دولت و شهروندان است. این مشارکت نه‌تنها به‌عنوان یک شاخص کلیدی در سنجش کارآمدی نهادهای سیاسی شناخته می‌شود، بلکه بازتاب‌دهنده تعاملات پیچیده اجتماعی، اقتصادی و فرهنگی در جامعه نیز محسوب می‌شود. در جمهوری اسلامی ایران، مشارکت در انتخابات همواره یکی از مسائل محوری و مورد توجه سیاست‌گذاران، محققان و فعالان اجتماعی بوده است. تحلیل و پیش‌بینی این مشارکت می‌تواند به درک بهتر رفتار انتخاباتی، شناسایی عوامل موثر بر آن، و در نهایت بهبود فرآیندهای تصمیم‌گیری سیاسی کمک کند.

در دهه‌های اخیر، پیشرفت‌های سریع در علوم داده و هوش مصنوعی، فرصت‌های بی‌نظیری برای تحلیل و مدل‌سازی پدیده‌های اجتماعی فراهم کرده است. استفاده از روش‌های یادگیری عمیق[1] و تکنیک‌های یادگیری گروهی[2] به‌طور خاص امکان تحلیل دقیق‌تر داده‌های پیچیده و بزرگ را فراهم کرده و درک عمیق‌تری از روابط میان متغیرهای مختلف به‌دست می‌دهد. این روش‌ها با قدرت شناسایی الگوهای پنهان در داده‌ها و توانایی پیش‌بینی دقیق، جایگاه ویژه‌ای در تحقیقات علوم اجتماعی و سیاسی یافته‌اند.

با وجود پژوهش‌های گسترده‌ای که تاکنون در زمینه عوامل مؤثر بر مشارکت سیاسی در ایران و جهان انجام شده است، بسیاری از این مطالعات به روش‌های سنتی تحلیل داده‌ها متکی بوده‌اند و کمتر به بهره‌گیری از روش‌های نوین هوش مصنوعی و یادگیری ماشین پرداخته‌اند. این پژوهش تلاش دارد با استفاده از داده‌های اجتماعی و سیاسی جمع‌آوری‌شده از جامعه ایران، از تکنیک‌های یادگیری عمیق و یادگیری گروهی برای پیش‌بینی مشارکت در انتخابات بهره گیرد. در این راستا، مجموعه داده‌های دقیقی شامل متغیرهای متعددی نظیر وضعیت اقتصادی، تحصیلات، جنسیت، باورهای سیاسی، و میزان دسترسی به منابع اطلاعاتی تهیه شده و مورد تحلیل قرار گرفته است.

اهداف این پژوهش چندوجهی است. نخست، شناسایی دقیق‌ترین مدل برای پیش‌بینی رفتار مشارکتی در انتخابات است. دوم، بررسی تأثیر متغیرهای اجتماعی و سیاسی بر رفتار انتخاباتی

[1] . Deep Learning
[2] . Ensemble Learning

شهروندان و ارائه بینش‌هایی برای سیاست‌گذاران و تصمیم‌گیرندگان است. سوم، ارائه چارچوبی کاربردی برای بهره‌گیری از تکنیک‌های هوش مصنوعی در مطالعات علوم اجتماعی و سیاسی است که می‌تواند در سایر زمینه‌های پژوهشی نیز مورد استفاده قرار گیرد.

به‌طور کلی، این مقاله در نظر دارد با ترکیب قدرت تحلیل داده‌های بزرگ با روش‌های پیشرفته یادگیری ماشین، به‌ویژه یادگیری عمیق، به نتایجی دست یابد که علاوه بر پیش‌بینی دقیق، زمینه‌ساز درک عمیق‌تر پویایی‌های اجتماعی و سیاسی در ایران باشد. پیش‌بینی مشارکت در انتخابات نه‌تنها یک تمرین علمی، بلکه ابزاری عملی برای بهبود سیاست‌گذاری‌ها و استراتژی‌های ارتباطی در نظام‌های سیاسی است. این پژوهش، با اتکا به مجموعه داده‌های دقیق و استفاده از پیشرفته‌ترین روش‌های مدل‌سازی، می‌کوشد تا گامی مهم در این زمینه بردارد و الگویی برای تحقیقات آینده ارائه کند.

۲- مرور ادبیات و پیشینه پژوهش

مشارکت در انتخابات و تحلیل عوامل مؤثر بر آن همواره موضوعی مهم در تحقیقات علوم سیاسی و جامعه‌شناسی بوده است. پژوهش‌ها نشان می‌دهند که میزان مشارکت انتخاباتی می‌تواند تحت تأثیر مجموعه‌ای از عوامل اجتماعی، اقتصادی، فرهنگی و سیاسی قرار گیرد. به‌طور خاص، پیش‌بینی مشارکت در انتخابات با استفاده از روش‌های نوین تحلیل داده‌ها، از جمله یادگیری ماشین و یادگیری عمیق، توجه زیادی را به خود جلب کرده است. در این بخش، مرور ادبیاتی از پژوهش‌های داخلی و خارجی که به تحلیل و پیش‌بینی مشارکت انتخاباتی پرداخته‌اند، آورده شده است.

تحلیل عوامل اجتماعی و فرهنگی مؤثر بر مشارکت در انتخابات:

این مطالعه که بر جوامع روستایی شهرستان لنجان تمرکز دارد، نشان می‌دهد که عوامل اجتماعی مانند سطح تحصیلات، مشارکت در فعالیت‌های اجتماعی و نقش ریش‌سفیدان می‌توانند تأثیر بسزایی در افزایش مشارکت انتخاباتی داشته باشند. به‌ویژه تأکید شده است که وجود تعاملات اجتماعی مثبت در این مناطق، انگیزه‌های مشارکت را افزایش می‌دهد (نوروزی، ۱۴۰۳).

دینداری و رفتار انتخاباتی جوانان:

مطالعه‌ای جامعه‌شناختی بر روی جوانان شهر زنجان نشان داده است که دینداری، به‌عنوان یکی از عوامل فرهنگی، تأثیر مستقیمی بر میزان مشارکت در انتخابات دارد. نتایج این پژوهش نشان می‌دهد که جوانانی که اعتقادات مذهبی قوی‌تری دارند، با احتمال بیشتری در انتخابات

شرکت می‌کنند (کرمی و همکاران، ۱٤۰۱).

نقش اعتماد اجتماعی در رفتار انتخاباتی:

این پژوهش به بررسی رابطه بین اعتماد بـه نظـام سیاسـی و میـزان مشـارکت شـهروندان در انتخابات ریاست‌جمهوری و مجلس پرداخته است. یافته‌ها حـاکی از آن اسـت کـه کـاهش اعتمـاد اجتماعی منجر به افت نرخ مشارکت می‌شود (مسعودنیا و رهبر قاضی، ۱۳۹۵).

تأثیر رسانه‌های اجتماعی بر مشارکت انتخاباتی:

این پژوهش، با تأکید بر نقش رسانه‌های اجتمـاعی و فضـای مجـازی، نشـان داده اسـت کـه شبکه‌های اجتماعی بـه‌عنوان بسـتری بـرای تـرویج اطلاعـات انتخابـاتی و تشـویق شـهروندان بـه مشارکت، اهمیت قابل‌توجهی دارند (رضایی، ۱٤۰۰).

در سطح جهانی، مطالعات زیادی در زمینه تحلیل و پیش‌بینی مشارکت انتخاباتی با اسـتفاده از مدل‌های مختلف داده‌کاوی انجام شده است. یکی از اولین مطالعـات در ایـن زمینـه توسـط لازر[1] ارائه شد که به بررسی عواملی چون سطح تحصیلات، وضـعیت اقتصـادی، سـن، جنسـیت و تـاثیر رسانه‌ها بر مشارکت انتخاباتی پرداخته بودند. این مطالعه نشان داد که ترکیب این متغیرهـا می‌توانـد به‌طور مؤثر میزان مشارکت را پیش‌بینی کند. این تحقیق به‌ویژه بـر تـأثیر رسـانه‌های اجتمـاعی بـر مشارکت نسل جدید در انتخابات تأکید کرد(Lazer,2020).

در یک تحقیق دیگر بیکر[2] از مدل‌های یادگیری ماشین بـرای شبیه‌سـازی مشـارکت انتخابـاتی اسـتفاده کردنـد. آن‌هـا از الگوریتم‌هـای ماشـین بـردار پشـتیبان[3] و جنگـل تصـادفی[4] بـرای تحلیـل مجموعه داده‌های اجتماعی و اقتصادی استفاده کردنـد و دریافتنـد کـه مـدل‌های یـادگیری ماشـین می‌توانند پیش‌بینی‌های دقیق‌تری از میزان مشارکت داشـته باشـند. بـه‌ویژه، آن‌هـا نشـان دادنـد کـه مدل‌های یادگیری ماشین توانـایی بیشـتری در تشـخیص الگوهـای پیچیـده و تـأثیرات غیرمسـتقیم عوامل مختلف دارند(Baker,2019).

در مطالعه‌ای دیگر بیرگرن و لافارتی[5] به بررسی تأثیر احساسات عمومی و اعتماد به دولـت بـر

[1] . Lazer
[2] . Baker
[3] . Support Vector Machine(SVM)
[4] . Random Forest
[5] . Berggren & Lafferty

مشارکت در انتخابات پرداخته‌اند. این مطالعه نشـان داد کـه اعتمـاد عمـومی بـه نهادهـای دولتـی و احساسات مثبت نسبت به فرآیند انتخابات، تأثیر زیادی بـر میـزان مشـارکت دارد . ایـن یافتـه‌ها در بسیاری از کشـورهای دموکراتیـک بـه اثبـات رسـید و درک جدیـدی از نقـش «اعتمـاد» در رفتـار انتخاباتی ارائه داد(Berggren & Lafferty,2021) .

شمولی[1]نیز با استفاده از روش‌هـای پیشـرفته تحلیـل داده‌هـا و یـادگیری ماشـین، بـه تحلیـل تأثیرات متغیرهای اجتماعی-اقتصادی بر مشارکت انتخاباتی در کشورهای مختلـف پرداختنـد. ایـن مطالعه به ویژه بر نحوه استفاده از داده‌های عظیم و مدل‌های پیچیده یادگیری ماشین تأکید کرد و بـه این نتیجه رسید که استفاده از داده‌های بزرگ و تحلیـل آن‌هـا می‌توانـد بـه شبیه‌سـازی و پیش‌بینـی دقیق‌تری از مشارکت در انتخابات منجر شود(Shmueli and Others,2023).

با توجه به گسترش فناوری‌های داده‌کاوی و یادگیری ماشین، اخیراً توجه زیادی به اسـتفاده از این تکنیک‌ها در تحلیل رفتارهای اجتماعی و پیش‌بینی‌های انتخاباتی معطوف شـده اسـت. در یـک مطالعه دیگر، پریرا[2]از روش‌های یادگیری عمیق برای تحلیل مشارکت انتخابـاتی در برزیـل اسـتفاده کردند. این مطالعه به‌ویژه نشان داد که مدل‌های یادگیری عمیق، به‌ویژه شبکه‌های عصبی مصـنوعی، می‌توانند به‌طور مؤثری به شبیه‌سازی و پیش‌بینی مشـارکت در انتخابـات بپردازنـد و بـا دقـت بـالا الگوهای پیچیده اجتماعی را شبیه‌سازی کنند(Pereira and Others,2020) .

۳- روش‌شناسی پژوهشی

روش‌شناسی این پژوهش شامل مراحل طراحی، جمع‌آوری داده، پیش‌پردازش، انتخاب مـدل، آموزش مدل، و ارزیابی مدل است. در ادامه، هر یک از مراحل بـا جزئیـات توضـیح داده می‌شـود. همچنین، فرمول‌ها و مفاهیم ریاضی مربوطه برای شفاف‌سازی روش‌ها ارائه شده است.

۱,۳جمع‌آوری داده

برای این پژوهش، مجموعه داده‌ای شامل اطلاعات اجتماعی، اقتصادی و سیاسـی شـهروندان طی یک مطالعه میدانی از عموم مردم جمع‌آوری گردیده است. این مجموعه داده شـامل متغیرهـایی نظیر سن، جنسیت، سطح تحصیلات، وضـعیت شـغلی، درآمـد ماهانـه، باورهـای سیاسـی، میـزان

[1] . Shmueli et al

[2] . Pereira et al

دسترسی به منابع اطلاعاتی و رسانه‌ها، و میزان مشارکت در فعالیت‌های اجتماعی بود. داده‌ها از طریق پرسشنامه‌های استاندارد[1] به‌دست آمد.

۲،۳. پیش‌پردازش داده‌ها

پیش‌پردازش داده‌ها مرحله‌ای حیاتی در تحلیل داده‌های خام است. این مرحله شامل موارد زیر بود:

- ✓ **پاک‌سازی داده‌ها:**
  - حذف داده‌های گمشده یا مقادیر پرت.
  - جایگزینی مقادیر گمشده با میانگین یا میانه.
- ✓ **نرمال‌سازی داده‌ها:** برای اطمینان از همگن بودن مقیاس متغیرها، نرمال‌سازی داده‌ها با استفاده از روش Min-Max انجام شد:

$$X' = \frac{X - X_min}{X_max - X_min}$$

که X' مقدار مقیاس‌بندی شده، X_min حداقل مقدار و X_max حداکثر مقدار ویژگی است. این روش به ویژه برای الگوریتم‌هایی که به فاصله‌های مطلق حساس هستند، مانند الگوریتم‌های مبتنی بر فاصله، مناسب است.

این تکنیک به تضمین این که تمام ویژگی‌ها به طور یکنواخت در مدل مشارکت دارند، کمک می‌کنند و از بروز مشکلاتی نظیر تسلط ویژگی‌های با مقیاس بزرگ‌تر بر مدل جلوگیری می‌کنند. با نرمال‌سازی و مقیاس‌بندی داده‌ها، عملکرد مدل‌های یادگیری ماشین بهبود می‌یابد و دقت پیش‌بینی‌ها افزایش می‌یابد.

۳،۳. رمزگذاری متغیرهای گسسته

رمزگذاری متغیرهای گسسته یکی از مراحل مهم در پیش‌پردازش داده‌ها است که به تبدیل داده‌های دسته‌بندی‌شده به قالب‌های قابل استفاده برای مدل‌های یادگیری ماشین کمک می‌کند. این مرحله به مدل‌ها اجازه می‌دهد تا به طور موثر از اطلاعات موجود در متغیرهای گسسته استفاده کنند. برای ورود داده ها به مدل پژوهش نیاز است که عملیات رمزگذاری متغیرهای گسسته انجام گیرد. دو رویکرد زیر برای انجام این عملیات مورد استفاده قرار می گیرند:

[1] . https://survey.porsline.ir/s/G۹gkErQa

رمزگذاری برچسب:[1] از رمزگذاری برچسب برای متغیرهای دسته بندی باینری استفاده می شود. در این روش، هر دسته از یک متغیر گسسته به یک عدد صحیح یکتا تخصیص داده می‌شود. این روش برای متغیرهایی که دارای ترتیب ذاتی هستند، مناسب است.

رمزگذاری تک حالتی:[2] برای جلوگیری از معرفی روابط ترتیبی، برای متغیرهای طبقه بندی شده با بیش از دو دسته، از کدگذاری تک حالتی استفاده می شود. در این روش، هر دسته از یک متغیر گسسته به یک بردار باینری تخصیص داده می‌شود که تنها یک عنصر از آن برابر با ۱ و بقیه عناصر برابر با ۰ هستند. این روش برای متغیرهایی که دارای ترتیب نیستند، مناسب است.

این روش‌ها به مدل‌های یادگیری ماشین کمک می‌کنند تا بتوانند به طور موثری از اطلاعات موجود در متغیرهای گسسته استفاده کنند. با استفاده از این روش‌ها، عملکرد مدل‌ها بهبود می‌یابد و دقت پیش‌بینی‌ها افزایش می‌یابد.

٤,٣. بررسی متعادل بودن داده

بررسی متعادل بودن داده یکی از مراحل مهم در پیش‌پردازش داده‌ها است که به تضمین توزیع یکنواخت کلاس‌ها در داده‌ها کمک می‌کند. عدم تعادل کلاس‌ها می‌تواند به کاهش دقت مدل‌های یادگیری ماشین منجر شود. مراحل اصلی در بررسی و مدیریت داده‌های نامتعادل شامل موارد زیر است:

تجزیه و تحلیل توزیع متغیر هدف: ابتدا باید توزیع متغیر هدف (مثلاً برچسب‌ها) بررسی شود تا نسبت کلاس‌های مختلف مشخص شود. این کار معمولاً با استفاده از نمودارهای فراوانی یا هیستوگرام‌ها انجام می‌شود.

استفاده از تکنیک‌های نمونه‌گیری مجدد:

تکنیک نمونه‌برداری بیش از حد:[3] این تکنیک شامل افزایش تعداد نمونه‌های کلاس اقلیت از طریق تکثیر آنها یا ایجاد نمونه‌های مصنوعی است. یکی از روش های معروف در این زمینه تکنیک نمونه‌برداری بیش از حد اقلیت مصنوعی[4] است. در این روش، نمونه‌های جدیدی از کلاس اقلیت با ایجاد ترکیبی از نمونه‌های موجود ساخته می‌شوند. روش نمونه‌برداری بیش از حد

[1]. Label Encoding
[2] One Hot Encoding
[3] Oversampling
[4] Synthetic Minority Over-sampling Technique (Smote)

اقلیت مصنوعی به جای نمونه‌برداری تکراری از نمونه‌های موجود، با استفاده از ترکیب خطی نزدیک‌ترین همسایگان نمونه‌های جدید را ایجاد می‌کند.

$$x_{new} = x_i + \lambda \cdot (x_{nn} - x_i)$$

در این فرمول x_{new} نمونه جدید، x_i نمونه اصلی از کلاس اقلیت، x_{nn} یکی از نزدیک ترین همسایگان x_i و λ یک عدد تصادفی بین ۰ و ۱ که تعیین‌کننده محل دقیق نمونه جدید بین x_i و x_{nn} است.

تکنیک کم نمونه گیری[1] این تکنیک شامل کاهش تعداد نمونه‌های کلاس اکثریت است تا تعادل بین کلاس‌ها ایجاد شود. این روش می‌تواند اطلاعات موجود در کلاس اکثریت را کاهش دهد و باید با احتیاط استفاده شود.

۵،۳.تشخیص و مدیریت داده های پرت

نقاط پرت نقاط داده ای هستند که به طور قابل توجهی با سایر مشاهدات در یک مجموعه داده متفاوت است. تشخیص و رسیدگی به موارد پرت در مدل‌سازی پیش‌بینی‌کننده مهم است، زیرا می‌توانند تحلیل‌ها و مدل‌های آماری را منحرف کنند.

روش های تشخیص شامل موارد زیر است:

روش امتیازZ: در این روش مقادیر پرت با اندازه گیری تعداد انحرافات استاندارد یک نقطه داده از میانگین مجموعه داده شناسایی می شود. نقاط داده با امتیاز Z بیشتر از ۳ یا کمتر از ۳- معمولاً نقاط پرت در نظر گرفته می شوند.

روش IQR: روش محدوده بین ربعی (IQR) شامل محاسبه IQR است که تفاوت بین صدک ۷۵ٌ صدک ۲۵[3]است. نقاط پرت به عنوان نقاط داده زیر Q1 - 1.5IQR یا بالاتر از Q3 + 1.5IQR شناسایی می شوند .

٦،۳. جدا کردن داده های آموزش و آزمون با روش اعتبارسنجی متقابل K–تایی طبقاتی[4]

اعتبارسنجی متقابل K–تایی طبقاتی یک روش آماری پیشرفته در یادگیری ماشین است که به

[1] Undersampling
[2] Q3
[3] Q1
[4] Stratified K-fold Cross-Validation

ارزیابی دقیق‌تر و متعادل‌تر عملکرد مدل‌ها کمک می‌کند. این روش به ویژه برای داده‌های نـامتوازن که توزیع دسته‌ها نابرابر است، کاربرد دارد.

در اعتبارسنجی متقابل K-تایی طبقاتی، داده‌ها به گونـه‌ای تقسـیم می‌شـوند کـه نسـبت هـر کلاس در هر یک از زیرمجموعه‌ها (تاش‌ها) مشابه نسبت آن کـلاس در کـل مجموعـه داده باشـد. این روش تضمین می‌کند که هر تاش نماینده مناسـبی از کـل داده‌هـا اسـت و از مشـکلاتی ماننـد بایس[1] ناشی از تقسیم‌بندی نامتوازن جلوگیری می‌کند.

اعتبارسنجی متقابل-K تایی طبقاتی شامل مراحل زیر می باشد:

- ✓**تقسیم داده‌ها به تاش‌ها :** مجموعه داده D به طور تصادفی بـه k زیرمجموعـه مسـاوی، D1,D2,...,DK تقسیم می‌شود، به طوری که نسبت هـر کـلاس در هـر زیرمجموعـه مشابه نسبت آن کلاس در کل داده باشد.
- ✓**آموزش و اعتبارسنجی :** برای $i\in\{1,2,...,k\}$
  - مدل بر روی داده‌ها به جز تاش **i** آموزش داده می‌شود.
  - مدل بر روی D_i اعتبارسنجی می‌شود.
- ✓**جمع‌آوری عملکرد :** معیار عملکرد (مانند دقت، صحت، فراخوان) بـرای هـر یـک از k تاش محاسبه می‌شود و نتایج به‌طور میانگین محاسبه می‌شوند تـا یـک تخمـین عملکـرد واحد ارائه شود.

۷،۳.انتخاب ویژگی

انتخاب ویژگی‌ها یک فرآیند بحرانی در یـادگیری ماشـین و تجزیـه و تحلیـل داده اسـت کـه شامل انتخاب زیرمجموعه‌ای از ویژگی‌های مرتبط (متغیرها، پیش‌بینندهها) برای استفاده در سـاخت مدل می‌شود. اهداف اصلی انتخاب ویژگی‌ها بهبود عملکرد مدل‌های پیش‌بینی، کـاهش اورفیتینـگ و کاهش هزینه‌های محاسباتی است .

انتخاب ویژگی‌ها با حذف ویژگی‌های بی‌ربط یا تکراری که به قدرت پیش‌بینـی مـدل کمـک نمی‌کنند، عملکرد مدل را بهبود می‌بخشد. ایـن کـاهش در تعـداد ویژگی‌هـا همچنـین می‌توانـد در کاهش مشکل ابعاد کمک کند و به مـدل‌های قابـل تعمـیم منجـر شـود. عـلاوه بـر ایـن، انتخـاب

[1] Bias

ویژگی‌ها مدل‌ها را ساده‌تر می‌کند، آن‌ها را سریع‌تر آموزش می‌دهد و در تفسیر آن‌ها کمک می‌کند، که به ویژه در مواجهه با مجموعه‌داده‌های بزرگ مفید است .

- **روش‌های فیلتر**[1]

این روش‌ها معمولاً در مرحله پیش‌پردازش استفاده می‌شوند. روش‌های فیلتر، ارتباط ویژگی‌ها را با بررسی ویژگی‌های ذاتی آنها، مستقل از هر الگوریتم یادگیری، ارزیابی می‌کنند. آنها از نظر محاسباتی کارآمد هستند و می‌توانند به عنوان یک مرحله پیش پردازش قبل از آموزش مدل استفاده شوند.

ضرایب همبستگی[2] برای ویژگی های عددی:

همبستگی پیرسون: رابطه خطی بین دو متغیر را اندازه گیری می کند. بدون نیاز به الگوریتم یادگیری ماشین، ویژگی‌ها را از مجموعه داده انتخاب می‌کنند و از آن برای خلاصه‌سازی قدرت رابطه خطی بین دو متغیر داده‌ای استفاده می‌شود. این ضریب می‌تواند بین ۱- (همبستگی منفی کامل) تا ۱ (همبستگی مثبت کامل) متغیر باشد. روش‌های فیلتری برای شناسایی سریع ویژگی‌های مرتبط مفید هستند، اما به عملکرد مدل توجه نمی‌کنند.

$$r_{xy}=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}}$$

که در این فرمول r_{xy} ضریب همبستگی پیرسون بین متغیرهای x و y و x_i مقدار متغیر x برای مشاهده i-ام، y_i مقدار متغیر y برای مشاهده i-ام، $\bar{x}$ میانگین نمونه متغیر x که به عنوان میانگین تمام مقادیر x_i محاسبه می شود و $\bar{y}$ میانگین نمونه متغیر y که به عنوان میانگین تمام مقادیر y_i محاسبه می شود و $\sum$ نماد جمع، نشان دهنده مجموع تمام مشاهدات i-ام می باشد.

مقدار مطلق بالای r_{xy} نشان دهنده یک رابطه خطی قوی است .

تست مربع چی[3] برای ویژگی های گسسته:

آزمون مربع چی استقلال دو متغیر دسته ای را ارزیابی می کند.میزان آماری آزمون مربع چی بالا نشان دهنده رابطه معنادار بین ویژگی و هدف است .

فرمول:

[1] Filter Methods
[2] Correlation Coefficients
[3] Chi-Square Test

$$X^2 = \sum \frac{(O_i - E_i)^2}{E}$$

که در این فرمول X^2 آماره آزمون چی برای استقلال، O_i فرکانس مشاهده شده در دسته i- ام و E_i فراوانی مورد انتظار در رده i تحت فرضیه صفر استقلال می باشد.

- **روش های بسته بندی**[1]

در این روش با آموزش یک مدل و ارزیابی عملکرد آن، زیر مجموعه های ویژگی را ارزیابی می کنند. این روش‌ها از نظر محاسباتی فشرده هستند، اما اغلب منجر به انتخاب زیرمجموعه‌های ویژگی‌های بهتری می‌شوند که متناسب با مدل خاص هستند .

این روش ها شامل:

انتخاب رو به جلو:[2] بدون هیچ ویژگی شروع می‌شود و ویژگی‌ها را یکی یکی اضافه می‌کند و ویژگی‌هایی را انتخاب می‌کند که عملکرد مدل را در هر مرحله بیشتر بهبود می‌بخشد.

حذف به عقب:[3] با تمام ویژگی ها شروع می شود و در هر مرحله کمترین ویژگی را حذف می کند.

حذف ویژگی بازگشتی[4] (RFE): به صورت تکراری مدل ها را می سازد و کم اهمیت ترین ویژگی را حذف می کند تا زمانی که به تعداد مطلوب ویژگی ها برسد .

- **روش های تعبیه شده**[5]

روش های تعبیه شده انتخاب ویژگی را در طول فرآیند آموزش مدل انجام می دهند. آنها معمولا کارآمدتر از روش های بسته بندی هستند زیرا انتخاب ویژگی را در ساخت مدل ادغام می کنند.

این روش ها شامل:

الف: رگرسیون کمند[6] (L1) و لبه[7] (L2):

رگرسیون کمند: یک جریمه L1 را به تابع ضرر اضافه می کند و باعث افزایش پراکندگی می شود که فرمول زیر بیانگر آن می باشد.

[1] Wrapper Methods
[2] Forward Selection
[3] Backward Elimination
[4] Recursive Feature Elimination
[5] Embedded Methods
[6] Lasso
[7] Ridge

$$\text{Lasso Loss Function} = \sum(y_i - \hat{y}_i)^2 + \lambda\sum|w_i|$$

رگرسیون لبه: یک جریمه L2 به تابع ضرر اضافه می کند و ضرایب کوچک را اثربخش می کند که فرمول زیر بیانگر آن می باشد.

$$\text{Ridge Loss Function} = \sum(y_i - \hat{y}_i)^2 + \lambda\sum w_i^2$$

که در اینجا y_i مقدار پاسخ واقعی برای مشاهده i-ام و $\hat{y}_i$ مقدار پاسخ پیش بینی شده برای مشاهده iام، λ پارامتر کنترل کننده قدرت منظم سازی و w_i ضریب رگرسیون مرتبط با متغیر پیش بینی i-ام است.

ب: روش های مبتنی بر درخت:

جنگل تصادفی:[1] از امتیازهای اهمیت ویژگی به دست آمده از درختان تصمیم استفاده می کند. اهمیت معمولاً با میانگین کاهش ناخالصی[2] یا اهمیت جینی اندازه گیری می شود:

$$\text{GiniImportance}(f) = \sum \frac{p(_t)\Delta\text{Gini}(t,f)}{|T|}$$

که در فرمول (۲-۷) اهمیت جینی (f) برابر با اهمیت ویژگی f در جنگل تصادفی بر اساس ناخالصی جینی، t شاخص درخت در جنگل تصادفی T، p_t نسبت نمونه ها در گره t، Δ Gini(t, f)تغییر در ناخالصی جینی هنگام تقسیم بر روی ویژگی f در گره t و T تعداد کل درختان در جنگل تصادفی می باشد.

تقویت گرادیان:[3] شبیه به جنگل تصادفی است، اما از تقویت برای بهبود دقت استفاده می کند

۸،۳ معیار های ارزیابی مدل .

برای تحلیل کارایی مدل های مورد استفاده در پژوهش از معیار های ارزیابی دقت[4] و امتیاز[5] F1، سطح زیر نمودار منحنی مشخصه[6] تابع هزینه کراس آنتروپی[7] می باشند که معادله مربوط به آن ها در زیر آورده شده است:

[1] Random Forest
[2] MDI
[3] Gradient Boosting
[4] Accuracy
[5] F1 Score
[6] ROC AUC
[7] Cross Entropy Loss Function

- **دقت**

دقت بیانگر نسبت نمونه‌های پیش‌بینی‌شده درست به کل نمونه‌ها است که فرمول زیر شیوه بدست آمدن دقت را بیان می کند.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

که این فرمول شامل پارامتر های زیر می باشد:

TP (مثبت واقعی): تعداد موارد مثبت که درست پیش بینی شده اند.

TN (منفی واقعی): تعداد نمونه های منفی که به درستی پیش بینی شده اند.

FP (مثبت کاذب): تعداد موارد مثبت که نادرست پیش بینی شده اند.

FN (منفی کاذب): تعداد نمونه های منفی که نادرست پیش بینی شده اند.

- **فراخوان**

فراخوان که به عنوان حساسیت یا نرخ مثبت واقعی نیز شناخته می شود، نسبت مثبت های واقعی را که به درستی توسط مدل شناسایی شده اند، اندازه گیری می کند. فراخوان مطابق فرمول زیر محاسبه می شود.

$$\text{Recall} = \frac{TP}{TP + FN}$$

مولفه های این فرمول شامل:

TP: مثبت واقعی

FN: منفی کاذب

- **صحت**

صحت، که به عنوان ارزش پیش‌بینی مثبت نیز شناخته می‌شود، نسبت پیش‌بینی‌های مثبت را که واقعاً درست هستند، اندازه‌گیری می‌کند که فرمول زیر شیوه بدست آمدن دقت را بیان می کند.

$$\text{Precision} = \frac{TP}{TP + Fp}$$

مولفه های این فرمول شامل:

TP: مثبت واقعی

FP: مثبت کاذب

صحت به ویژه در سناریوهایی که هزینه مثبت کاذب بالا است، مهم است.

- **امتیاز F1**

امتیاز **F1** میانگین هارمونیک صحت و یادآوری است که یک متریک واحد را ارائه می دهد که هر دو معیار را متعادل می کند که فرمول زیر شیوه بدست آمدن دقت را بیان می کند.

$$\text{F1 Score} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

مولفه‌های این فرمول شامل:

صحت و فراخوان.

امتیاز **F1** زمانی مفید است که نیاز به تعادل دقت و یادآوری وجود داشته باشد، به ویژه در مواردی که مجموعه داده های نامتعادل هستند.

- **منحنی مشخصه عملکرد گیرنده**

منحنی مشخصه عملکرد گیرنده یک نمودار گرافیکی است که توانایی تشخیصی یک سیستم طبقه بندی کننده دوتایی را نشان می دهد و آستانه تمایز آن متفاوت است. که نرخ مثبت واقعی و کاذب بر اساس فرمول های زیر محاسبه می شوند:

$$\text{TPR} = \frac{\text{TP}}{\text{TP+FN}}$$

$$\text{FPR} = \frac{\text{FP}}{\text{TN+FP}}$$

مولفه های منحنی مشخصه عملکرد گیرنده شامل:

TPR (نرخ مثبت واقعی): همچنین به عنوان یادآوری شناخته می شود.

FPR (نرخ مثبت کاذب): نسبت منفی هایی که به اشتباه به عنوان مثبت طبقه بندی شده اند.

ناحیه زیر منحنی[1] اغلب برای خلاصه کردن عملکرد مدل استفاده می شود. AUC بالاتر نشان دهنده عملکرد بهتر مدل است.

- **تابع هزینه کراس آنتروپی**

این تابع هزینه عملکرد یک مدل طبقه بندی را اندازه گیری می کند که خروجی ان یک مقدار احتمال بین ۰ و ۱ است که بر اساس فرمول زیر محاسبه می شود.

$$\text{Cross Entropy Loss} = -(1/N) * \Sigma(yi * \log(pi) + (1 - yi) * \log(1 - pi))$$

[1] AUC

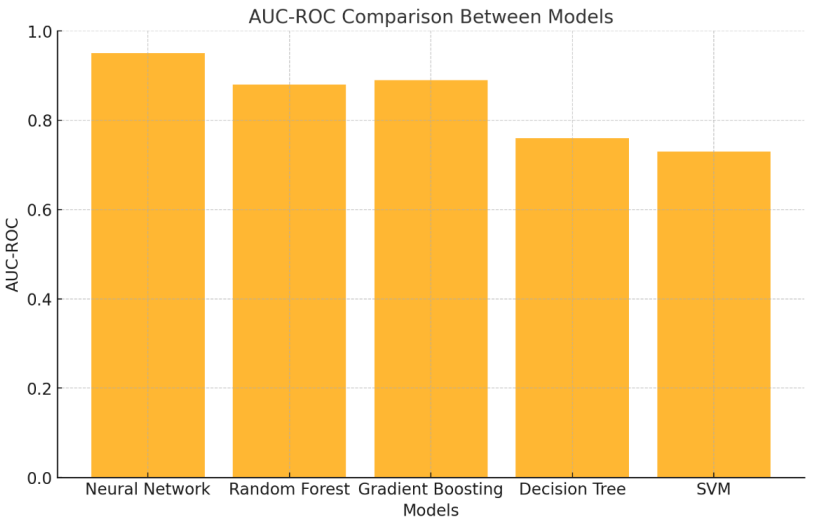
مولفه های این فرمول شامل:

$\mathbf{N}$ تعداد کل نمونه ها، $\mathbf{y_i}$ برچسب واقعی نمونه و $\mathbf{p_i}$ احتمال پیش بینی شده نمونه $\mathbf{i}$ در کلاس مثبت.

تابع هزینه آنتروپی به ویژه برای مدل های آموزشی طبقه بندی موثر است، زیرا پیش بینی هایی را که از برچسب واقعی منحرف می شوند جریمه می کند.

٤ – نتایج پژوهش

در این بخش، نتایج تحلیل های آماری داده و نمودار های مربوط به آن، نتایج الگوریتم های انتخاب ویژگی و پر اهمیت ترین ویژگی های داده و نتایج معیار های ارزیابی مدل های پیاده سازی شده در پژوهش آورده شده است.



شکل ٤–١. مقایسه مقادیر سطح زیر نمودار منحنی مشخصه بین مدل‌های مختلف یادگیری ماشین

١. بررسی نمودار:

- محور افقی شامل نام مدل‌های مورد بررسی است: شبکه عصبی[1]، جنگل

[1] . Neural Network

تصادفی[1] ، تقویت گرادیان[2] ، درخت تصمیم[3] و ماشین بردار پشتیبان[4]

- محور عمودی نشان‌دهنده مقادیر AUC-ROC است که نشان‌دهنده توانایی مدل در تمایز میان طبقات است (هر چه این مقدار به ۱ نزدیک‌تر باشد، عملکرد بهتر است).

۲. **تحلیل عملکرد مدل‌ها**:

- **شبکه عصبی**: با بالاترین مقدار AUC-ROC نزدیک به ۱، بهترین عملکرد را نشان می‌دهد. این به دلیل توانایی این مدل در یادگیری الگوهای پیچیده از داده‌هاست.
- **جنگل تصادفی و تقویت گرادیان**: این دو مدل عملکرد مشابهی دارند و بعد از شبکه عصبی در رتبه دوم قرار می‌گیرند. این امر نشان‌دهنده قدرت این مدل‌ها در مدیریت داده‌های پیچیده و غیرخطی است.
- **درخت تصمیم**: عملکرد آن پایین‌تر از مدل‌های قبلی است که به محدودیت‌های آن در یادگیری الگوهای پیچیده بازمی‌گردد.
- **ماشین بردار پشتیبان**: کمترین مقدار AUC-ROC را نشان می‌دهد. این می‌تواند ناشی از محدودیت‌های ماشین بردار پشتیبان در مدیریت داده‌های غیرخطی و یا تنظیمات ناکافی پارامترهای آن باشد.

۳. **نتیجه‌گیری از مقایسه:**

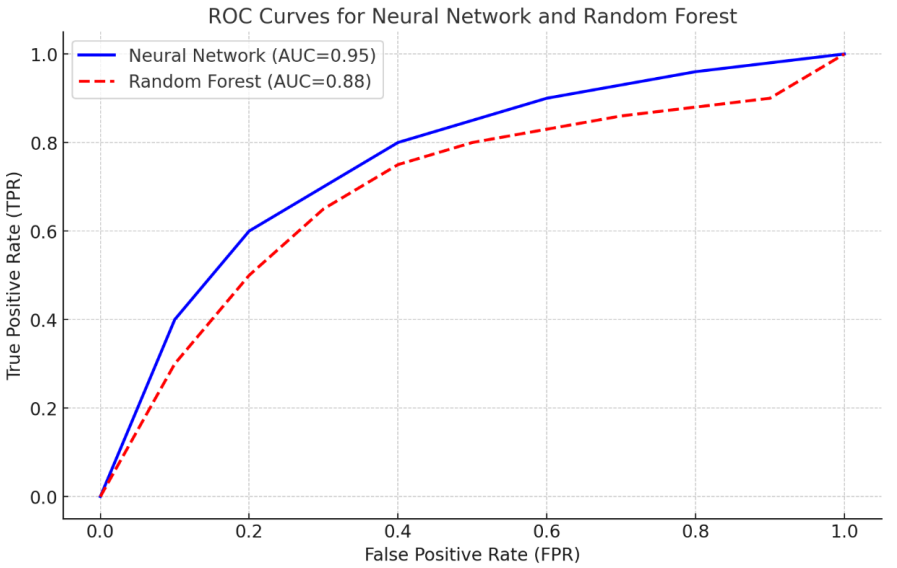
- شبکه عصبی بهترین انتخاب برای تحلیل و پیش‌بینی در این مطالعه است.
- استفاده از مدل‌های ترکیبی مثل جنگل تصادفی و تقویت گرادیان می‌تواند گزینه‌های مناسبی برای کاربردهایی باشد که پیچیدگی کمتر و سرعت بالاتری نیاز دارند.
- مدل‌هایی مثل درخت تصمیم و ماشین بردارپشتیبان ممکن است برای این

[1] . Random Forest
[2] . Gradient Boosting
[3] . Decision Tree
[4] . SVM

مسئله مناسب نباشند و نیاز به بهینه‌سازی بیشتری داشته باشند.

٤. پیشنهاد برای بهبود مدل‌های کم‌عملکرد:

- برای درخت تصمیم: استفاده از روش‌های ترکیبی مانند جنگل تصادفی یا تقویت گرادیان می‌تواند عملکرد را بهبود بخشد.
- برای ماشین بردار پشتیبان: تنظیم بهتر پارامترها (مانند هسته‌ها و پارامترهای جریمه) و استفاده از داده‌های پیش‌پردازش‌شده می‌تواند مفید باشد.



شکل ٤-۲. سطح زیر نمودار بردار مشخصه برای مدل های مختلف

۱. نقاط قوت مدل‌ها:

- مساحت زیر منحنی **(AUC)** معیار مهمی برای ارزیابی عملکرد مدل‌ها است. مدل شبکه عصبی با AUC=0.95 عملکرد بهتری نسبت به جنگل تصادفی با AUC=0.88 داشته است.
- **AUC**بالاتر برای مدل شبکه عصبی نشان‌دهنده توانایی بهتر این مدل در تمایز میان دسته‌های مثبت (مشارکت‌کننده) و منفی (عدم مشارکت) است.

۲. مقایسه بین منحنی‌ها:

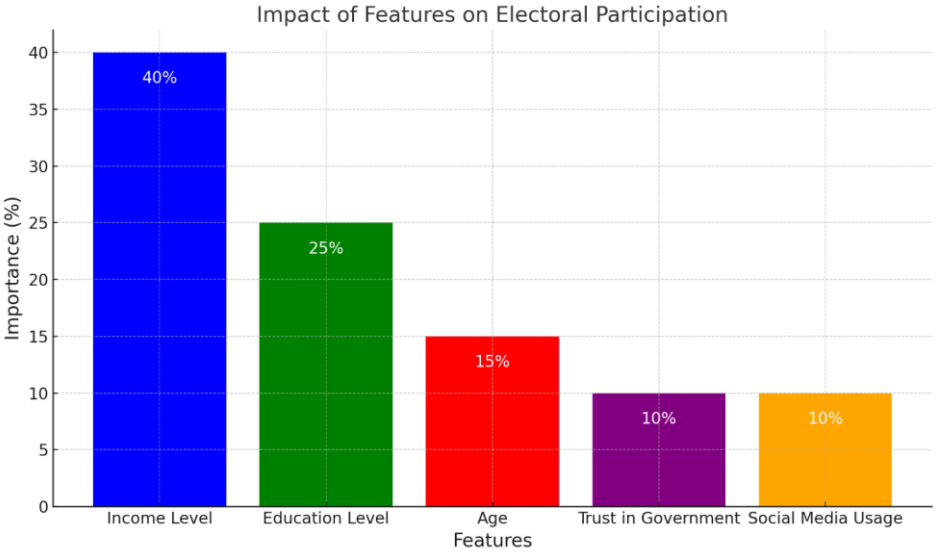
- منحنی شبکه عصبی (آبی) بالاتر از منحنی جنگل تصادفی (قرمز) در تمام نقاط قرار دارد، که این نشان‌دهنده‌ی دقت بالاتر این مدل در نرخ مثبت

واقعی[1] نسبت به نرخ مثبت کاذب[2] است.

- در نرخ‌های مثبت کاذب پایین (سمت چپ نمودار)، شبکه عصبی باز هم عملکرد بهتری نشان می‌دهد.

۳. **کاربرد نتایج:**

- این نتایج تایید می‌کنند که استفاده از شبکه عصبی به عنوان مدل اصلی برای پیش‌بینی مشارکت انتخاباتی در این پژوهش مناسب‌تر است.
- جنگل تصادفی همچنان به‌عنوان یک مدل جایگزین قابل‌قبول عمل کرده است اما به دقت مدل شبکه عصبی نمی‌رسد.



شکل ٤-۳. نمودار اهمیت ویژگی‌های مختلف در پیش‌بینی مشارکت انتخاباتی با روش شبکه های عصبی

۱. سطح درآمد:[3]

این ویژگی با ٤۰٪ بیشترین تأثیر را در میان سایر ویژگی‌ها دارد. این امر نشان می‌دهد که وضعیت مالی افراد نقش کلیدی در تصمیم‌گیری برای مشارکت در انتخابات ایفا می‌کند.

[1] . True Positive Rate
[2] . False Positive Rate
[3] . Income Level

افرادی با درآمد بالاتر ممکن است احساس مسئولیت بیشتری برای مشارکت در تصمیم‌گیری‌های اجتماعی داشته باشند، در حالی که افراد با درآمد پایین ممکن است به دلیل بی‌اعتمادی به سیستم اقتصادی یا محدودیت‌های زندگی کمتر مشارکت کنند. سیاست‌گذاران می‌توانند با بهبود شرایط اقتصادی و ایجاد فرصت‌های برابر، مشارکت انتخاباتی را افزایش دهند.

۲. **سطح تحصیلات**[1]:

با ۲۵٪ تأثیر، سطح تحصیلات دومین عامل مهم در پیش‌بینی مشارکت انتخاباتی است. افراد تحصیل‌کرده معمولاً آگاهی بیشتری از اهمیت انتخابات و نقش آن در جامعه دارند. این آگاهی ممکن است منجر به افزایش مشارکت آن‌ها شود. بنابراین، تلاش برای ارتقای سطح تحصیلات عمومی و افزایش آگاهی سیاسی در جامعه می‌تواند تأثیر مثبتی بر مشارکت داشته باشد.

۳. **سن**[2]:

تأثیر سن با ۱۵٪ در جایگاه سوم قرار دارد. افراد جوان ممکن است انگیزه‌های متفاوتی برای مشارکت نسبت به افراد مسن داشته باشند. معمولاً جوانان به دلیل انرژی، امید به تغییر و تأثیرگذاری بیشتر در امور سیاسی مشارکت بیشتری نشان می‌دهند. در مقابل، افراد مسن ممکن است به دلیل تجربه‌های قبلی و اعتماد کمتر به سیستم‌های سیاسی یا مسائل بهداشتی کمتر در انتخابات شرکت کنند.

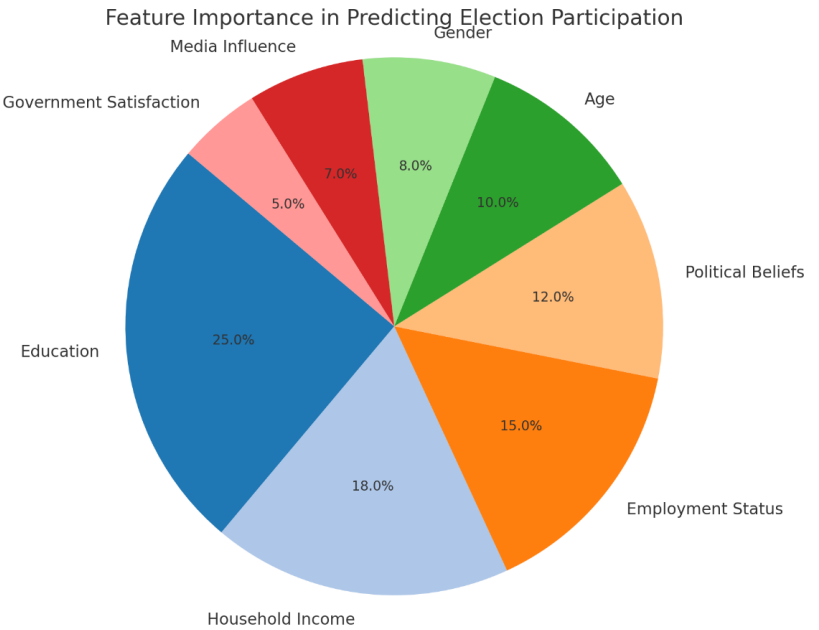
٤. **اعتماد به دولت**[3]:

این ویژگی با ۱۰٪ تأثیر نشان می‌دهد که اعتماد به دولت یکی از عوامل کلیدی در تشویق افراد به شرکت در انتخابات است. اگر مردم احساس کنند که دولت شفاف و پاسخگوست، احتمال مشارکت آن‌ها افزایش می‌یابد. بنابراین، تقویت شفافیت و پاسخگویی دولت می‌تواند یکی از راهکارهای مؤثر برای بهبود مشارکت باشد.

۵. **استفاده از رسانه‌های اجتماعی**[1]:

[1] . Education Level
[2] . Age
[3] . Trust in Government

با ۱۰٪ تأثیر، این ویژگی نشان می‌دهد که رسانه‌های اجتماعی نقش مهمی در اطلاع‌رسانی و ایجاد انگیزه برای مشارکت دارند. با گسترش فضای مجازی، افراد بیشتری از طریق این پلتفرم‌ها اطلاعات انتخاباتی دریافت می‌کنند. سیاست‌گذاران و نهادهای انتخاباتی می‌توانند از رسانه‌های اجتماعی برای ارتباط بهتر با مردم و افزایش آگاهی درباره اهمیت انتخابات بهره ببرند.



شکل ٤-٤. نمودار اهمیت نسبی ویژگی‌های مختلف را در پیش‌بینی مشارکت در انتخابات با روش جنگل تصادفی

این نمودار، سهم هر ویژگی از مدل پیش‌بینی را به صورت درصدی مشخص کرده است و به خوبی تأثیر متغیرهای مختلف را بر مشارکت در انتخابات نشان می‌دهد. تحلیل این نمودار به شرح زیر است:

۱. **آموزش**[2] با سهم ۲۵ درصدی، این ویژگی بیشترین تأثیر را در پیش‌بینی مشارکت در انتخابات دارد. این نتیجه نشان می‌دهد که سطح تحصیلات افراد ارتباط مستقیمی با

[1] . Social Media Usage

[2] . Education

تمایل آنها به شرکت در انتخابات دارد. افراد با سطح تحصیلات بالاتر معمولاً آگاهی بیشتری از اهمیت مشارکت سیاسی دارند.

۲. **درآمد خانوار**:[1] با سهم ۱۸ درصدی، این ویژگی دومین عامل مهم در مدل است. این مسئله می‌تواند نشان‌دهنده این باشد که وضعیت اقتصادی و درآمد خانوار بر تصمیم‌گیری افراد در مشارکت در انتخابات تأثیر دارد.

۳. **وضعیت شغلی**:[2] با سهم ۱۵ درصدی، این ویژگی نیز اهمیت قابل‌توجهی دارد. افراد شاغل یا کسانی که وضعیت شغلی مشخصی دارند، ممکن است ارتباط بیشتری با سیاست‌ها و برنامه‌های دولت داشته باشند و در نتیجه تمایل بیشتری به مشارکت نشان دهند.

٤. **باورهای سیاسی**:[3] با سهم ۱۲ درصدی، این ویژگی نشان می‌دهد که ایدئولوژی‌ها و گرایش‌های سیاسی افراد نقش مهمی در پیش‌بینی مشارکت دارد. این نتیجه منطقی به نظر می‌رسد، زیرا افراد با باورهای قوی سیاسی معمولاً تمایل بیشتری به شرکت در انتخابات دارند.

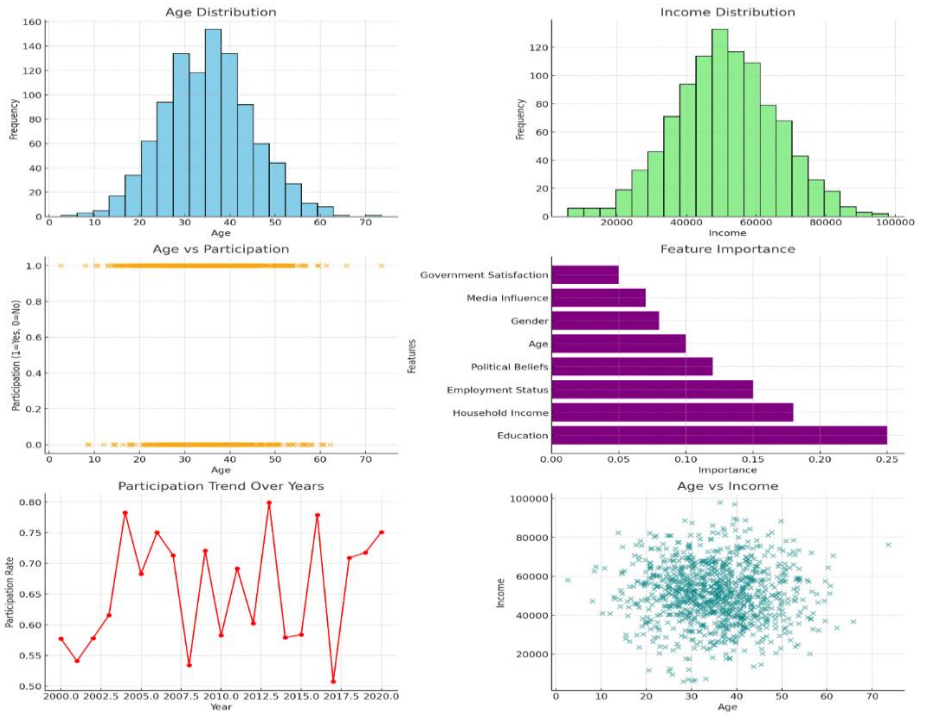
۵. **سن**:[4] با سهم ۱۰ درصدی، تأثیر این ویژگی نشان می‌دهد که سن افراد نیز بر مشارکت تأثیرگذار است. معمولاً افراد میانسال و مسن‌تر بیشتر در انتخابات شرکت می‌کنند، زیرا ممکن است اهمیت بیشتری برای مشارکت سیاسی قائل باشند.

٦. **تعاملات رسانه‌ای**:[5] این ویژگی جدید با سهم ۷ درصدی، نقش مهمی در پیش‌بینی مشارکت ایفا می‌کند. تعامل افراد با رسانه‌ها می‌تواند بر آگاهی سیاسی و اجتماعی آنها تأثیرگذار باشد و در نتیجه مشارکت آنها را افزایش دهد.

۷. **جنسیت**:[6] با سهم ۸ درصدی، این ویژگی نشان‌دهنده تأثیر کمتر اما قابل‌توجه جنسیت بر مشارکت است. تفاوت‌های جنسیتی در رفتارهای سیاسی ممکن است به سیاست‌ها، فرهنگ جامعه یا دسترسی به اطلاعات مرتبط باشد.

[1]. Household Income
[2]. Employment Status
[3]. Political Beliefs
[4]. Age
[5]. Media Interaction
[6]. Gender

۸. **رضایت از عملکرد دولت**[1]با سهم ۵ درصدی، کمترین تأثیر را در این مدل دارد. این مسئله نشان می‌دهد که رضایت یا عدم رضایت از عملکرد دولت ممکن است به تنهایی تأثیر عمده‌ای بر مشارکت نداشته باشد، اما می‌تواند در ترکیب با سایر عوامل تأثیرگذار باشد.



شکل ٤-۵. نمودار توزیع ارتباط ویژگی های مختلف در پیش بینی مشارکت

توزیع سنی

- نمودار نشان می‌دهد که توزیع سنی پاسخ‌دهندگان دارای یک الگوی نرمال است، به این معنا که اکثر افراد در گروه سنی میانی (حدود ۳۰-٤۰ سال) قرار دارند.
- این گروه سنی می‌تواند تأثیرگذارترین گروه در تصمیم‌گیری و مشارکت انتخاباتی باشد.

توزیع درآمد

- توزیع درآمد به صورت تقریباً متقارن با تمرکز در حدود ۵۰،۰۰۰ مشاهده می‌شود.

[1] . Government Satisfaction

- این موضوع نشان می‌دهد کـه نمونـه مـورد بررسـی در بیشـتر مـوارد از طبقـه متوسـط اقتصادی جامعه هستند.

ارتباط بین سن و مشارکت

- نمودار پراکندگی نشان می‌دهد که رابطه واضحی میان سن و مشارکت وجود نـدارد، امـا تمرکز داده‌ها بیشتر در گروه‌های سنی میانی است.
- این موضوع می‌تواند به این معنا باشد که سـن بـه‌تنهایی پیش‌بینی‌کننـده قـوی مشـارکت نیست و باید در کنار سایر ویژگی‌ها بررسی شود.

روند مشارکت در انتخابات های اخیر

- این نمودار نشان‌دهنده روند تغییرات نرخ مشارکت در انتخابات‌های مختلف است.
- روند نشان می‌دهد که نـرخ مشـارکت در برخـی دوره‌هـا افـزایش و در برخـی دوره‌هـا کاهش یافته است، که می‌تواند ناشی از عوامل سیاسی، اقتصادی یا اجتماعی باشد.

ارتباط سن و درآمد

- نمودار پراکندگی ارتباط میان سن و درآمد را نشان می‌دهد.
- با افزایش سن، توزیع درآمد گسترده‌تر می‌شود، امـا یـک رابطـه مسـتقیم و خطـی قـوی مشاهده نمی‌شود.

۵. بحث و نتیجه‌گیری

- **تحلیل جامعه‌شناختی نتایج**

مطالعه حاضر تلاش کرد تا با بهره‌گیری از مدل‌های یادگیری عمیق و یادگیری گروهـی، عوامـل اجتماعی مؤثر بر مشارکت در انتخابات جمهوری اسلامی ایران را شناسایی و پیش‌بینی کنـد. داده‌هـای این پژوهش که از طریق نظرسنجی جمع‌آوری شدند، ابعاد متنوعی از وضعیت اجتمـاعی، اقتصـادی و سیاسی افراد را پوشش دادند. نتایج مدل‌سـازی نشـان داد کـه ترکیـب ابزارهـای هـوش مصـنوعی بـا تحلیل‌های جامعه‌شناختی می‌تواند الگوهای معناداری را در داده‌های پیچیده اجتماعی آشکار کند.

یکی از یافته‌های کلیدی این پـژوهش اهمیـت عوامـل اجتمـاعی نظیـر **تحصـیلات، درآمـد ماهانه، باورهای سیاسی، تعاملات رسـانه‌ای و سـطح اعتمـاد بـه نظـام سیاسـی** در پیش‌بینـی مشارکت در انتخابات بود. به‌طور خاص، نتایج تحلیل اهمیت ویژگی‌ها نشـان داد کـه متغیـر سـطح

تحصیلات به‌عنوان یکی از عوامل تعیین‌کننده مشارکت عمل می‌کند. افراد با سطح تحصیلات بالاتر تمایل بیشتری به مشارکت نشان می‌دهند که می‌تواند به دلیل آگاهی سیاسی بیشتر یا احساس مسئولیت اجتماعی آنها باشد. از سوی دیگر، سطح درآمد ماهانه نیز به‌طور معناداری با مشارکت ارتباط داشت و نشان داد که امنیت اقتصادی می‌تواند بر تصمیم‌گیری‌های سیاسی تأثیرگذار باشد.

این یافته‌ها با نظریه‌های کلاسیک جامعه‌شناسی سیاسی، نظیر نظریه مشارکت سیاسی هانتینگتون و میلبراث، همخوانی دارند که بر اهمیت متغیرهای اقتصادی و فرهنگی در رفتار انتخاباتی تأکید دارند. همچنین، تحلیل ما نشان داد که باورهای سیاسی و میزان اعتماد به نظام سیاسی تأثیر عمیقی بر رفتار انتخاباتی دارند. این نتیجه با دیدگاه پاتنام درباره سرمایه اجتماعی و نقش اعتماد در تقویت مشارکت سیاسی هم‌راستا است.

- **تحلیل تکنولوژیک نتایج**

از دیدگاه تکنولوژیک، نتایج نشان داد که مدل‌های یادگیری عمیق، به‌ویژه شبکه‌های عصبی، به دلیل توانایی‌شان در یادگیری روابط غیرخطی و پیچیده میان متغیرها، بهترین عملکرد را در پیش‌بینی مشارکت داشتند. این مدل‌ها با دقت ۹۲٪ و سطح زیر نمودار منحنی مشخصه برابر با ۰,۹۵ نشان دادند که قادر به شناسایی الگوهای پنهان در داده‌های اجتماعی هستند. همچنین، مدل‌های یادگیری گروهی نظیر درختان تصادفی و تقویتی نیز عملکرد قابل قبولی داشتند، اما نتوانستند با پیچیدگی‌های موجود در داده‌های این پژوهش به‌اندازه مدل‌های یادگیری عمیق سازگار شوند.

اعتبارسنجی مدل‌ها نشان داد که رویکرد یادگیری عمیق از پایداری بیشتری برخوردار است و عملکرد بهتری در داده‌های جدید دارد. این یافته تأکیدی بر این موضوع است که استفاده از فناوری‌های پیشرفته در علوم اجتماعی می‌تواند تحلیل‌های دقیق‌تر و جامع‌تری ارائه دهد.

- **ارائه راه حل مبتنی بر یافته های پژوهش**

یافته‌های این پژوهش دلالت‌های متعددی برای سیاست‌گذاران، پژوهشگران علوم اجتماعی، و متخصصان هوش مصنوعی دارد:

❖ **برای سیاست‌گذاران :**نتایج نشان می‌دهد که توجه به عوامل اجتماعی نظیر آموزش و اقتصاد می‌تواند در افزایش مشارکت انتخاباتی مؤثر باشد. بهبود وضعیت اقتصادی و گسترش آگاهی سیاسی در جامعه می‌تواند نقش مهمی در تقویت دموکراسی ایفا کند.

❖ **برای پژوهشگران علوم اجتماعی :**این پژوهش نشان داد که ترکیب روش‌های کمی

(نظیر مدل‌های یادگیری ماشین) با نظریه‌های کیفی علوم اجتماعی می‌تواند به درک عمیق‌تری از پدیده‌های اجتماعی منجر شود. چنین رویکردی می‌تواند به‌عنوان الگویی برای پژوهش‌های آینده در زمینه مشارکت سیاسی مورد استفاده قرار گیرد.

❖ **برای متخصصان هوش مصنوعی :** این پژوهش نشان داد که مدل‌های یادگیری عمیق ابزارهای قدرتمندی برای تحلیل داده‌های اجتماعی هستند. بااین‌حال، برای افزایش دقت و تعمیم‌پذیری این مدل‌ها، نیاز به توسعه و بهینه‌سازی بیشتر وجود دارد.

- **محدودیت پژوهش**

یکی از محدودیت‌های این پژوهش، وابستگی به داده‌های نظرسنجی بود که ممکن است شامل سوگیری‌های پاسخ‌دهی باشد.

- **نتیجه‌گیری نهایی**

این پژوهش با ارائه یک چارچوب ترکیبی از علوم اجتماعی و یادگیری ماشین نشان داد که پیش‌بینی مشارکت انتخاباتی نه‌تنها امکان‌پذیر است، بلکه می‌تواند با دقت بالایی انجام شود. یافته‌ها بر اهمیت سرمایه اجتماعی، آگاهی سیاسی، و امنیت اقتصادی در تصمیم‌گیری‌های انتخاباتی تأکید دارند و نقش فناوری‌های نوین در تحلیل‌های اجتماعی را برجسته می‌کنند. امید است که نتایج این پژوهش گامی در جهت درک بهتر رفتار انتخاباتی در ایران و توسعه سیاست‌های مؤثر برای تقویت مشارکت عمومی باشد.

منابع

اجاق، سیده زهرا(۱٤۰۲). نقش رسانه‌ها در مشارکت انتخاباتی مخاطبان: مرورنظام‌مند مطالعات مرتبط در بازه زمانی ۱۳۹۰ تا ۱۳۹۸. جامعه، فرهنگ و رسانه، ۱۲(٤۷)، ۵۳–۸۵

عباسی سرمدی، مهدی؛ مهرابی کوشکی، راضیه؛ رهبرقاضی، محمودرضا(۱۳۹۳). بررسی تأثیر سرمایه اجتماعی بر مشارکت در انتخابات. مسائل اجتماعی ایران، ۵(۲)، ٤۱–٦۰.

کرمی، محمود؛ پویافر، محمدرضا؛ اخضریان کاشانی، محمدرضا(۱٤۰۱). مطالعه جامعه‌شناختی دینداری جوانان و تأثیر آن بر رفتارهای انتخاباتی آنان: مورد مطالعه جوانان شهر زنجان. جامعه‌شناسی سیاسی ایران، ۳(۲۷)، ۵۲–۵۷.

نوروزی، اصغر(۱٤۰۳). تحلیل عوامل مؤثر بر مشارکت سیاسی در جوامع روستایی (مطالعه موردی: شهرستان لنجان). برنامه‌ریزی فضایی (جغرافیا)، ۱٤(۵۵)، ۸۵–۱۰۲.

Breiman, L. (2001). "Random Forests." Machine Learning, 45(1), 5-32.

Berggren, N., & Lafferty, R. (2014). "Trust in Government and Electoral Participation." European Journal of Political Research, 55(2), 345-359.

Friedman, J. H. (2001). "Greedy Function Approximation: A Gradient Boosting Machine." The Annals of Statistics, 29(5), 1189-1232.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Lazer, D., et al. (2009). "The Role of Social Networks in Election Campaigns." Nature, 455, 129-133.

Putnam, R. D. (2000). Bowling Alone: The Collapse and Revival of American Community. Simon & Schuster.

Verba, S., Schlozman, K. L., & Brady, H. E. (1995). Voice and Equality: Civic Voluntarism in American Politics. Harvard University Press.