

ارزیابی تکنیک‌های هوش محاسباتی در پیش‌بینی سرانه تولید پسماند (مطالعه موردی: استان هرمزگان)

محمد مینوسپهر^۱، محمدرضا علیزاده^۱ و ناصر طالب‌بیدختی^{۲*}

^۱ کارشناسی ارشد مهندسی عمران محیط‌زیست، دانشکده مهندسی، دانشگاه شیراز

^۲ استاد گروه راه، ساختمان و محیط‌زیست، دانشکده مهندسی، دانشگاه شیراز

(دریافت: ۱۳۹۵/۵/۱۶، پذیرش: ۱۳۹۵/۱۱/۹، نشر آنلاین: ۱۳۹۵/۱۱/۱۰)

چکیده

شناخت کمیت پسماندهای یک شهر یا منطقه، لازمه برنامه‌ریزی در زمینه مدیریت پسماند است. روش دستیابی به کمیت پسماندها دانستن سرانه یا نرخ تولید آن است. در خصوص پیش‌بینی مقادیر سرانه تولید پسماند تا کنون در اکثر مدل‌های تدوین شده از داده‌های سری زمانی مربوط به منطقه مورد مطالعه استفاده شده است. اما در شرایطی که چنین داده‌هایی موجود نباشد استفاده از سیستم‌های هوشمند پیش‌بینی نظیر تکنیک‌های یادگیری ماشین که بر اساس داده‌های اندازه‌گیری شده در یک سال تدوین شوند، بسیار مؤثر خواهند بود. از آنجا که داده‌های زمان‌مندی جهت مقادیر سرانه تولید پسماند مناطق جمعیتی ساحلی جنوب ایران جهت طراحی اصولی سیستم مدیریت پسماند وجود نداشته است، در این مطالعه با در نظر گرفتن پارامترهای ارتفاع از سطح دریا، جمعیت، درجه شهری و تناوب جمع‌آوری پسماند، توانایی روش‌های هوشمند MLP، SVM و M5P در پیش‌بینی سرانه تولید پسماند شهرها و روستاهای ساحلی استان هرمزگان بکار گرفته شده و با هم مقایسه شده است. نتایج حاصله نشان دهنده این است که روش M5P با مقدار مجذور میانگین مربعات خطا (RMSE) $55/34$ (gr/d) و میانگین قدر مطلق خطای نسبی (MARE) $6/26$ درصد، بهترین عملکرد را نسبت به سایر روش‌ها دارد.

کلیدواژه‌ها: پیش‌بینی سرانه تولید پسماند، شبکه عصبی پرسپترون، ماشین بردار پشتیبان، مدل درخت تصمیم.

۱- مقدمه

دهه گذشته در زمینه پیش‌بینی نرخ تولید پسماند توسط روش‌های رگرسیونی (Jenkins, ۱۹۹۳ و Hockett, ۱۹۹۵ و Daskalopoulos, ۱۹۹۸ و Bogner, ۲۰۰۳ و Bach, ۲۰۰۴)، آنالیز سری‌های زمانی (Chang, ۱۹۹۷ و Chen, ۲۰۰۰ و Navarro-Esbrí, ۲۰۰۲) و نیز روش‌های دینامیکی (Dyson, ۲۰۰۵ و Karavezyris, ۲۰۰۲) تلاش‌های زیادی صورت گرفته است. همچنین اخیراً علاقه فزاینده‌ای در توسعه تئوری سیستم‌های هوشمند مانند شبکه‌های عصبی در پیش‌بینی نرخ تولید پسماند در زمینه‌های مختلف مورد توجه محققین قرار گرفت (Dong, ۲۰۰۳ و Karaca, ۲۰۰۶ و Bayar, ۲۰۰۹). برای نمونه Jahandideh و همکاران (۲۰۰۹) از مدل‌های شبکه عصبی در پیش‌بینی سرانه تولید پسماند عفونی استفاده کردند. همچنین امکان پیش‌بینی میزان نرخ تولید پسماندهای شهری با استفاده از شبکه‌های عصبی مصنوعی با در نظر گرفتن شاخص‌های

آگاهی از کمیت پسماند تولیدی، یکی از اصول بنیادی در تمام بخش‌های مدیریت پسماند می‌باشد. شناخت کمیت پسماندها تأثیر اساسی در حجم سرمایه‌گذاری، انتخاب نوع مخازن ذخیره‌سازی و ماشین‌آلات، تعداد پرسنل مورد نیاز، ظرفیت محل‌های دفن و ... خواهد داشت. دلیل اصلی برآورد کمیت پسماند تولید شده، به دست آوردن داده‌هایی است که می‌توان از آن‌ها برای توسعه و اجرای مؤثر برنامه‌های مدیریت پسماند استفاده کرد (Tchobanoglous و همکاران، ۱۹۹۳). در بیشتر مواقع، به دلیل بودجه ناکافی، تجهیزات اندازه‌گیری محدود و فقدان سیستم مدیریتی مناسب، با نبود داده‌های ثبت شده زمان‌مند از مقادیر پسماند مواجه هستیم (Chang و Dyson, ۲۰۰۵). بنابر این دستیابی به یک پیش‌بینی دقیق با در نظر گرفتن نرخ تولید پسماند یک چالش مهم محسوب می‌شود. در

* نویسنده مسئول؛ شماره تماس: ۰۹۱۷۱۱۷۹۴۰۷

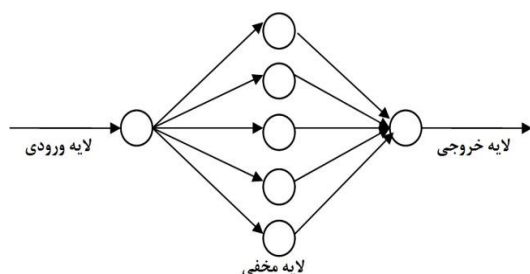
مصنوعی، قابل فهم تر هستند.

۲- مواد و روش‌ها

در این مطالعه توانایی سیستم‌های هوشمند شبکه عصبی پرسپترون چند لایه (MLP^5)، ماشین بردار پشتیبان (SVM) و درخت تصمیم (M5P) در پیش‌بینی سرانه تولید پسماند مورد توجه قرار گرفته است که در ادامه شرح داده می‌شوند.

۲-۱- شبکه عصبی پرسپترون چند لایه (MLP)

شبکه‌های عصبی، سیستم‌های هوشمندی هستند که با پردازش داده‌های مشاهده‌ای، دانش و یا قانون نهفته بین داده‌های ورودی و خروجی را به ساختار شبکه منتقل می‌کنند. به عبارت ساده‌تر، این سیستم‌های مبتنی بر هوش محاسباتی سعی می‌کنند ساختار عصبی مغز بشر را مدل‌سازی کنند. در سال ۱۹۴۳، زمانی که Warren McCulloch و Walter Pitts نشان دادند که شبکه‌های عصبی می‌توانند هر تابع حسابی را محاسبه نمایند، دیدگاه علمی شبکه‌های عصبی مصنوعی آغاز شد. کار این افراد با ارائه مکانیسم شرط‌گذاری^۶ جهت یادگیری نرون‌های بیولوژیکی توسط Donald O. Hebb ادامه یافت. نخستین کاربرد عملی شبکه‌های عصبی با معرفی شبکه پرسپترون توسط Frank Rosenblatt در سال ۱۹۵۸، آغاز شد. Rosenblatt و همکارانش شبکه‌ای ساختند که قادر بود الگوها را از هم شناسایی نماید. شبکه عصبی MLP مجموعه‌ای از نرون‌ها است که در لایه‌های مختلفی پشت سر هم قرار گرفته‌اند. مقادیر ورودی پس از ضرب در وزن‌های موجود در گذرگاه‌های بین لایه‌ها به نرون بعدی رسیده و در آن جا با هم جمع می‌شوند و پس از عبور از تابع شبکه مربوطه خروجی نرون‌ها را تشکیل می‌دهند (شکل ۱). در پایان خروجی به دست آمده با خروجی مورد نظر مقایسه شده و خطای به دست آمده جهت اصلاح وزن‌های شبکه به کار می‌رود، این امر اصطلاحاً آموزش شبکه عصبی نامیده می‌شود.



شکل ۱- ساختار شبکه عصبی پرسپترون

عمومی توسعه مانند GDP^۱ (Antanasijević و همکاران، ۲۰۱۳) و نیز شاخص‌های اجتماعی- اقتصادی (Batinić و همکاران، ۲۰۱۱) مورد بررسی قرار گرفته‌اند. Abdoli و همکاران (۲۰۱۲)، از شبکه عصبی مصنوعی و داده‌های سری زمانی ماهانه سال‌های ۲۰۱۰-۲۰۰۰ شهر مشهد، جهت پیش‌بینی طولانی مدت نرخ تولید پسماند در این شهر استفاده کرده‌اند. علاوه بر این برای انتخاب بهترین سناریوی ورودی داده و نیز تبدیل داده‌های ورودی به داده‌های مفیدتر، روش‌های پیش پردازش مورد توجه قرار گرفت که می‌توان در این مورد به مطالعات Noori و همکاران (۲۰۱۰)، با استفاده از روش‌های آنالیز مؤلفه‌های اصلی (PCA^2) و آزمون گاما^۳ در شبکه‌های عصبی برای پیش‌بینی تولید هفتگی پسماند جامد در مشهد اشاره کرد. بعد از تدوین مدل شبکه عصبی مصنوعی مناسب با ۱۳ ورودی، از آنالیز مؤلفه‌های اصلی و آزمون گاما برای کاهش تعداد ورودی‌ها استفاده شد. در نهایت با مقایسه بین نتایج شبکه عصبی به تنهایی و با وجود آنالیز مؤلفه‌های اصلی و آزمون گاما انجام شده نشان داد که آنالیز مؤلفه‌های اصلی و آزمون گاما تأثیر چشمگیری در نتایج داشته و آن‌ها را بهبود بخشیده است. این دو مدل تعداد ورودی‌های شبکه را به ترتیب از ۱۳ به ۷ و ۵ کاهش دادند.

پیش از این Noori و همکاران (۲۰۰۹)، پیش‌بینی تولید هفتگی پسماند با استفاده از ترکیبی از سیستم تبدیل موج فازی و تبدیل موج شبکه عصبی مصنوعی در تهران را مورد بررسی قرار داده بودند. نتایج به دست آمده در این مطالعه تأثیر مثبت پیش پردازش با استفاده از تبدیل موجی و افزایش قابل توجه در دقت دو مدل را نشان می‌دهد. از طرفی محدودیت‌هایی مانند آموزش بیش از حد شبکه، افتادن در دام بهینه‌های محلی و تدوین بهترین ساختار شبکه و مشکلاتی از این دست کارایی مدل‌های شبکه عصبی را محدودتر می‌کند. با ارائه روش‌هایی چون ماشین‌های بردار پشتیبان (SVM^4) توسط Vapnik (۱۹۹۵)، کارایی این روش‌ها نسبت به شبکه عصبی بسیار بهبود پیدا کرده و در پیدا کردن هم‌زمان بهینه عمومی عملکرد بهتری را نشان می‌دادند.

در همین راستا Abbasi و همکاران (۲۰۱۳)، با ترکیب روش ماشین‌های بردار پشتیبان به عنوان یک روش هوشمند و نیز روش مربعات جزئی (PLS^5) به عنوان یک ابزار انتخاب ویژگی به پیش‌بینی مقادیر هفتگی پسماند تولیدی در تهران پرداختند. از طرفی مدل درخت تصمیم‌گیری ($M5P^6$) که اولین بار توسط Quinlan (۱۹۹۲)، پیشنهاد شد از جمله مدل‌های نوین شبیه‌سازی می‌باشند که نسبت به مدل‌های درخت رگرسیونی از دقت بیشتری برخوردار بوده و هم‌زمان قوانین تعیین شده توسط آن نسبت به سایر مدل‌های هوشمند نظیر شبکه‌های عصبی

5. partial least square
6. M5P Model Trees
7. Multi-Layer Perceptron
8. Conditioning

1. Gross Domestic Product
2. Principal Component Analysis
3. Gamma Test (GT)
4. Support vector machines

راستای عمود بر مرز تصمیم‌گیری تا جایی که ممکن است حداکثر شود. در ماشین بردار پشتیبان به دو طریق می‌توان مجموعه نقاط را از هم جدا کرد: ۱- خطی، ۲- غیر خطی (AVCI، ۲۰۰۹). در حالت خطی، ماشین بردار پشتیبان با در نظر گرفتن مجموعه آموزشی ابرصفحه بهینه با حداکثر حاشیه را از طریق حل مسأله بهینه‌سازی زیر پیدا می‌کند:

با توجه به انتخاب مقادیر w و b ابرصفحه‌های جداکننده زیادی وجود دارد. در روش ماشین بردار پشتیبان تلاش می‌شود تا بهترین صفحه جداکننده دو کلاس به دست آید. ثابت می‌شود که صفحه بهینه دارای بیشترین فاصله بین مرزها است. رابطه $w \cdot x + b = 1$ بیانگر مرز ناحیه $y = 1$ و $w \cdot x + b = -1$ بیانگر مرز ناحیه $y = -1$ و نشان دهنده صفحه جدا کننده مرزهای دو ناحیه است.

$$w \cdot x + b \geq 1 \quad y = 1 \quad (1)$$

$$w \cdot x + b \leq -1 \quad y = -1 \quad (2)$$

در نتیجه:

$$y \sin g(w \cdot x + b) \geq 1 \quad (3)$$

حد فاصل بین دو ناحیه توسط تفاضل مرزهای ۲ ناحیه زیر به دست می‌آید:

$$w \cdot x + b = 1 \quad (4)$$

$$w \cdot z + b = -1 \quad (5)$$

در نتیجه داریم:

$$w \cdot (x - z) = 2, w / \|w\| \cdot (x - z) = 2 / \|w\| \quad (6)$$

حال، جهت به دست آوردن مرز بهینه باید فاصله بین دو ناحیه کمینه شود، لذا $\|w\|$ ناحیه بیشینه شود؛ یعنی مقدار داریم:

$$\text{Min} 1 / 2 \|w\|^2; y_k ((w \cdot x_k) + b) \geq 1 \quad (7)$$

$$k = 1, 2, \dots, m \quad (8)$$

حل این مسأله بهینه‌سازی کار مشکلی است، برای ساده‌تر کردن آن از روش ضرایب لاگرانژ استفاده و مقادیر بهینه w و b به دست آورده می‌شود.

$$\text{Max}(\text{Min}L(w, b, \alpha)) \quad (9)$$

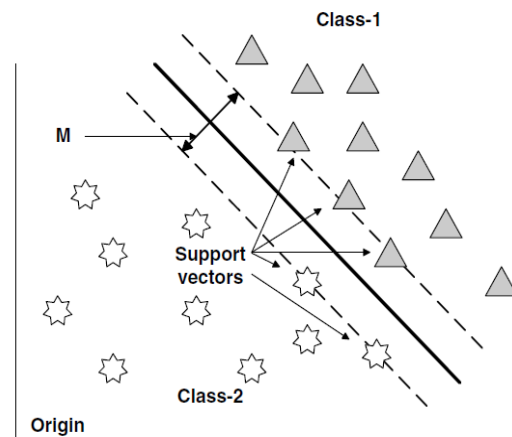
$$L(w, b, \alpha) = 1/2 \|w\|^2 - \sum_{k=1}^m \alpha_k (y_k ((w \cdot x_k) + b) - 1) \quad (10)$$

دوگان روابط فوق به صورت زیر بیان می‌شود:

این شبکه بر مبنای الگوریتم پس انتشار خطا آموزش می‌بیند. بدین ترتیب که خروجی‌های واقعی با خروجی‌های دلخواه مقایسه می‌شوند و وزن‌ها به وسیله الگوریتم پس انتشار، به صورت تحت نظارت تنظیم می‌گردند تا الگوی مناسب به وجود آید (Kalogirou، ۲۰۰۳).

۲-۲- ماشین‌های بردار پشتیبان (SVM)

ماشین بردار پشتیبان یک طبقه‌بندی کننده دوتایی است که با استفاده از نگاشت داده‌ها از فضای ورودی اصلی به فضایی بعد بالاتر برای جداسازی آن‌ها عمل می‌کند. این مدل ابرصفحه‌ای را جستجو می‌کند که فاصله‌اش با داده‌های دو کلاس ماکزیمم است. در این روش سعی بر آن است تا جهت به دست آوردن مرز کلاس‌ها، سیستمی با ظرفیت کمینه و یا به بیان بهتر سیستمی با حداقل پیچیدگی پیاده‌سازی شود. در نتیجه ماشین بردار پشتیبان می‌تواند با استفاده از داده‌های آموزشی کمتر نسبت به روش‌های رقیب، مرزهای سیستم را با دقت مناسبی تخمین بزند، بدون آن که تعمیم‌پذیری سیستم را مخدوش نماید. در واقع هدف اصلی طبقه‌بندی کننده ماشین بردار پشتیبان دستیابی به یک تابع $f(x)$ است که این تابع تعیین کننده ابرصفحه است. این ابرصفحه دو کلاس را به طور بهینه از هم جدا می‌کند که در شکل (۲) نشان داده شده است. در این شکل M حاشیه است یعنی فاصله ابرصفحه از نزدیک‌ترین نقاط هر دو کلاس (Huang و Wanmg، ۲۰۰۹).



شکل ۲- دسته‌بندی به دو کلاس توسط ماشین بردار پشتیبان

Vapnik (۱۹۹۵) برای محاسبه مرز جداکننده ۲ کلاس کاملاً جدا از هم، روش حاشیه بهینه را پیشنهاد کرد. در واقع باید یک مرز خطی را بین ۲ کلاس به گونه‌ای محاسبه کرد تا تمام داده‌هایی که به کلاس ۱+ تعلق دارند در یک طرف مرز قرار گیرند و تمام داده‌هایی که متعلق به کلاس ۱- هستند در طرف دیگر مرز واقع شوند. مرز تصمیم‌گیری بایستی به گونه‌ای باشد که فاصله نزدیک‌ترین نمونه‌های آموزشی هر دو کلاس از یکدیگر در

مدل رگرسیون خطی چندگانه ساخته می‌شود. در مرحله بعد مدل‌های رگرسیون خطی با کنار گذاشتن صفاتی که حذف آنها باعث کاهش میانگین خطا می‌گردد، ساده‌سازی می‌شوند. بعد از این ساده‌سازی، هر زیر درخت برای انجام عملیات هرس کردن مورد بررسی قرار می‌گیرد. اگر خطای تخمین زده شده برای مدل در ریشه زیر درخت، کوچک‌تر یا مساوی میانگین خطای درخت باشد، زیر درخت هرس می‌شود. مدل نهایی در یک برگ از ترکیب مدل به دست آمده در آن برگ با مدل‌های موجود در مسیر ریشه تا برگ مربوطه به دست می‌آید. کاهش انحراف معیار در این روش با استفاده از رابطه زیر محاسبه می‌گردد.

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i) \quad (15)$$

در رابطه فوق T شامل نمونه‌هایی است که به گره رسیده‌اند و T_i ها مجموعه‌هایی هستند که از تقسیم کردن گره بر اساس صفت انتخابی به دست آمده‌اند. sd انحراف معیار بوده و SDR نیز کاهش انحراف معیار می‌باشد (Wang و Witten، ۱۹۹۷). برای ارزیابی توانایی عمومی مدل M5P در پیش‌بینی، از تکنیک n-Fold Cross-Validation که از روش‌های رایج برای ارزیابی الگوریتم‌های یادگیری پایگاه داده است، استفاده شده است. این روش مجموعه داده‌ها را به طور تصادفی به n قسمت مساوی که Fold نامیده می‌شود تقسیم می‌کند. برای ساخت مدل از $n-1$ قسمت داده‌ها برای آموزش و از یک قسمت باقی‌مانده که برای ساخت مدل استفاده نشده است، برای تست مدل استفاده می‌کند. در این شیوه ساخت و تست مدل n بار تکرار شده و در هر بار داده‌های آموزش و تست تغییر می‌کنند تا هر n گروه به عنوان گروه تست به کار روند (Mingers، ۱۹۸۹).

۳- محدودده مورد مطالعه و روش کار

ناحیه مورد مطالعه در برگزیده ۲۲۰ شهر و روستای ساحلی استان هرمزگان به طول تقریبی ۹۰۰ کیلومتر واقع در جنوب ایران است. این ناحیه از کوشکنار در غربی‌ترین نقطه استان که هم‌جوار با خلیج فارس می‌باشد شروع و به بندر جاسک در شرقی‌ترین نقطه استان که هم‌جوار با دریای عمان می‌باشد ختم می‌شود (شکل (۳)). به منظور طراحی اصولی سیستم مدیریت پسماند شهری و روستایی مناطق ساحلی هرمزگان با افق ده ساله و به دلیل نبود آمار و اطلاعات از سرانه تولید پسماند، در این مطالعه ابتدا با هماهنگی انجام شده با مسئولان خدمات شهری و دهیاران محدوده مورد مطالعه، کل پسماند تولیدی تمامی شهرها و روستاها به صورت هفت روز متوالی در وسط هر یک از چهار فصل سال ۱۳۹۲، اندازه‌گیری و پس از آنالیز و بررسی داده‌ها، سرانه تولید پسماند شهرها و روستاها به دست آمده است. با توجه به هزینه‌های بالای آماربرداری و مشکلات اجرایی، نمونه‌برداری

$$\frac{\partial}{\partial b} L(w, b, \alpha) = 0 \rightarrow \sum \alpha_k y_k = 0 \quad (11)$$

$$\frac{\partial}{\partial w} L(w, b, \alpha) = 0 \rightarrow w = \sum \alpha_k y_k x_k \quad (12)$$

حال اگر مقدار w را در تابع $f(x) = \sin g(w.x + b)$ جاگذاری کنیم، رابطه زیر به دست می‌آید:

$$f(x) = \text{sign}(\sum \alpha_k y_k (x \cdot x_k) + b) \quad (13)$$

ضرایب αk در رابطه فوق، ضرایب لاگرانژ نامیده می‌شوند. به ازای بسیاری از αk ها، ضرایب xk برابر صفر است و تنها ضرایب محدودی غیر صفر است. نقاطی که به ازای آن‌ها ضرایب αk مخالف صفر هستند، بردارهای پشتیبان نامیده می‌شود و همین نقاط، تعیین کننده مرز جدا کننده هستند. هرچه مقدار αk کوچک‌تر باشد تأثیر xk در مرز جدا کننده کمتر است. در نقاط بردار پشتیبان αk ، مخالف صفر است. در این صورت در این نقاط داریم:

$$yk((w.x) + b) = 1 \quad (14)$$

یعنی بردارهای پشتیبان روی مرز قرار دارند. الگوریتم بالا برای یافتن مرز دو کلاس کاملاً جدا از هم بود که حاصل آن یک مرز خطی است.

۲-۳- الگوریتم مدل درختی (M5P)

از مدل‌های نوین شبیه‌سازی می‌توان به مدل درخت تصمیم-گیری اشاره کرد. مدل M5P از جمله مدل‌های درخت تصمیم-گیری می‌باشد که نسبت به مدل‌های درخت رگرسیونی از دقت بیشتری برخوردار بوده و هم‌زمان قوانین تعیین شده توسط آن نسبت به سایر مدل‌های هوشمند نظیر شبکه‌های عصبی مصنوعی، قابل فهم‌تر هستند. از جمله ویژگی‌های بارز مدل‌های M5P ارائه روابط رگرسیونی و معادلات ریاضی کاربردی بسیار ساده و قابل فهم و در عین حال با دقت مناسب جهت تخمین و پیش‌بینی می‌باشد (Quinlan، ۱۹۹۲). در ابتدا الگوریتم مدل درختی M5P فضای نمونه را به زیرمجموعه‌های درونی کوچک‌تر تقسیم می‌کند. این تقسیم کردن طوری صورت می‌پذیرد که تغییرات زیرمجموعه‌های درونی کمینه شود. این تغییرات به وسیله انحراف معیار مقادیری که از ریشه به هر گره رسیده‌اند، با محاسبه کاهش خطای مورد انتظار که از آزمایش هر صفت در آن گره حاصل می‌شود، اندازه‌گیری می‌گردد. صفتی که میزان کاهش خطای مورد انتظار را بیشینه کند، به عنوان صفت جدا کننده انتخاب می‌شود. فرآیند جداسازی در صورتی که مقادیر خروجی تمامی نمونه‌هایی که به گره مورد نظر می‌رسند تغییرات ناچیزی داشته باشند و یا تعداد کمی از نمونه‌ها باقی بمانند، متوقف می‌شود. پس از ایجاد درخت برای هر گره داخلی یک

سال تأسیس شهرداری، وسعت محدوده شهری، جمعیت، درآمد و امکانات بهداشت و آموزش، ارزیابی و اعلام می‌شود. پارامترهای ارتفاع از سطح دریا و جمعیت از داده‌های آماری رسمی ارائه شده توسط مرکز آمار ایران (سرشماری ۱۳۹۰) استخراج شده‌اند. جهت تناوب جمع‌آوری پسماند از نتایج مطالعات طرح‌های جامع مدیریت پسماند شهری و روستایی محدوده مورد مطالعه استفاده شده است. درجه شهریت نیز بر اساس آخرین آمار ارائه شده توسط سازمان شهرداری‌ها و دهیاری‌های وزارت کشور، منظور شده است.

به منظور ارزیابی عملکرد مدل‌های به کار رفته در پیش‌بینی مقادیر سرانه تولید پسماند از شاخص‌های آماری مجذور میانگین مربعات خطا (RMSE)^۹، میانگین خطای مطلق (MAE)^{۱۰}، ضریب همبستگی خطی (CC)^{۱۱}، میانگین قدر مطلق خطای نسبی (MARE)^{۱۲} استفاده شده است.

$$RMSE = \sqrt{\frac{1}{n} \sum (x_i - y_i)^2} \quad (16)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (17)$$

$$CC = \frac{\left(\frac{1}{n}\right) \left[(x_i - \mu_x)^T (y_i - \mu_y) \right]}{\sqrt{(1/n)(x_i - \mu_x)^2} \sqrt{(1/n)(y_i - \mu_y)^2}} \quad (18)$$

$$MARE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right| \times 100 \quad (19)$$

در تمامی فرمول‌های فوق x_i مقادیر واقعی داده‌ها و y_i مقادیر پیش‌بینی شده توسط مدل‌های هوشمند می‌باشد، n تعداد داده‌ها، μ_x و μ_y میانگین مقادیر x_i و y_i هستند.

مکرر در تمامی طول سال در منطقه امکان‌پذیر نبوده است، اما حتی‌الامکان سعی شده است آماربرداری در میانه‌های فصول مختلف سال برای در نظر گرفتن اثر تغییرات فصول بر میزان پسماند تولیدی، صورت گیرد (مینوسپهر، ۱۳۹۲). سپس با استفاده از روش‌های SVM، MLP و M5P، مقادیر سرانه تولید پسماند برای سال‌های ۱۳۹۳ تا ۱۴۰۲ پیش‌بینی شده است. پارامترهای بسیار زیادی وجود دارند که می‌توانند در سرانه تولید پسماند در یک منطقه تأثیرگذار باشند، اما شاید کمی کردن و در نظر گرفتن همه این پارامترها و ارزیابی میزان اثرات هر کدام با توجه به عدم امکان ثبت و ضبط این داده‌ها در دوره‌های متوالی، کاری دشوار و غیرممکن باشد. با تمام این محدودیت‌ها، در ساختار تدوین شده در این مقاله ابتدا پارامترهای مختلفی از جمله حداقل دما، حداکثر دما، میزان بارندگی، رطوبت، درجه شهریت، ارتفاع از سطح دریا و تناوب جمع‌آوری که داده‌های ثبت شده در یک دوره زمانی مناسب برای آن‌ها وجود داشته است، در نظر گرفته شده‌اند. سپس با انجام آنالیز حساسیت بر روی این پارامترها، مؤثرترین پارامترها به عنوان ورودی در تدوین ساختار نهایی مدل‌های هوش محاسباتی استفاده شده است. بدیهی است در صورت امکان آماربرداری و اندازه‌گیری سایر پارامترهای مهم و تأثیرگذار، می‌توان مدل‌هایی با دقت بالاتر و جامع‌تر ارائه کرد. پارامترهای ارتفاع از سطح دریا، جمعیت، درجه شهریت و تناوب جمع‌آوری پسماند به عنوان داده‌های ورودی جهت پیش‌بینی مقادیر سرانه تولید پسماند مورد استفاده قرار گرفته‌اند. درجه شهریت یا درجه شهرداری هر منطقه‌ای به خدماتی که با ایجاد زیرساخت‌ها برای رفاه شهروندان ایجاد می‌شود، بستگی دارد. به عبارتی درجه شهرداری بیانی از شاخص توسعه برای مناطق مختلف می‌باشد که توسط سازمان شهرداری‌ها و دهیاری‌های وزارت کشور برای هر شهر بر اساس زیر شاخص‌های



شکل ۳- محدوده مورد مطالعه

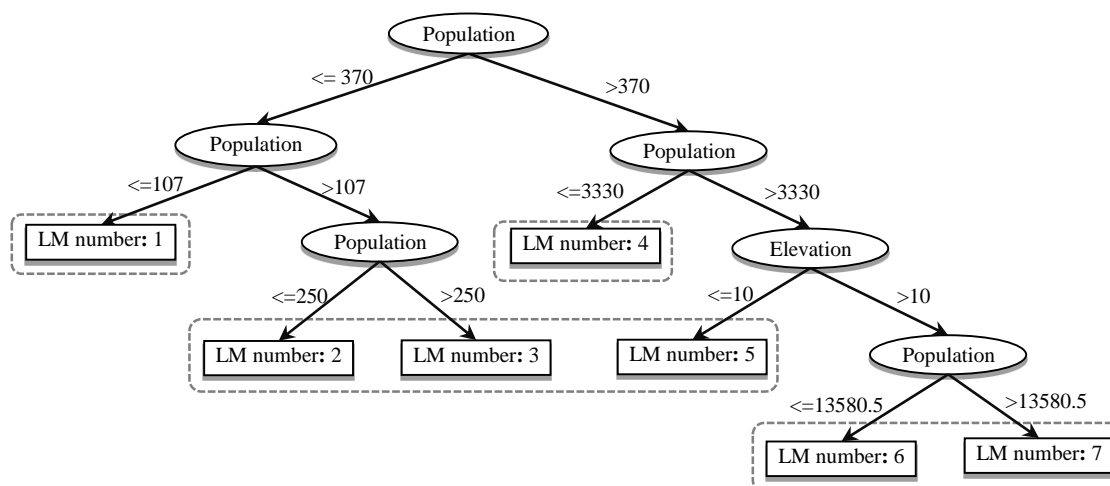
11. Correlation Coefficient
12. Mean Absolute Relative Error

9. Root Mean Square Error
10. Mean Absolute Error

۴- نتایج

به کمترین مقدار خطا رسیده است. در شکل (۴) ساختار درختی قواعد حاصل شده از مدل M5P ارائه است. با توجه به شکل، تعداد ۷ قاعده در پیش‌بینی مقادیر سرانه تولید پسماند توسط مدل حاصل شده که در جدول (۱) ارائه شده است. همان طور که در جدول (۱) مشاهده می‌شود، مدل M5P توانسته است با استفاده از روابط ساده ریاضی، ارتباط و پیچیدگی پارامترهای مؤثر بر تولید سرانه پسماند به عنوان ورودی مدل، شامل جمعیت (P)، درجه شهریت (D)، ارتفاع از سطح دریا (E) و تناوب جمع‌آوری پسماند (F) را به خوبی تحلیل کند. با توجه به ساختار درختی قواعد حاصل شده از مدل M5P در پیش‌بینی سرانه تولید پسماند در شکل (۴)، ابتدا بر اساس مقادیر پارامتر جمعیت (P) یکی از قوانین حاصل شده (LM) انتخاب می‌شود، سپس با توجه به ضرایب حاصل شده از مدل در جدول (۱)، مقدار سرانه تولید پسماند طبق معادله حاکمه به دست خواهد آمد.

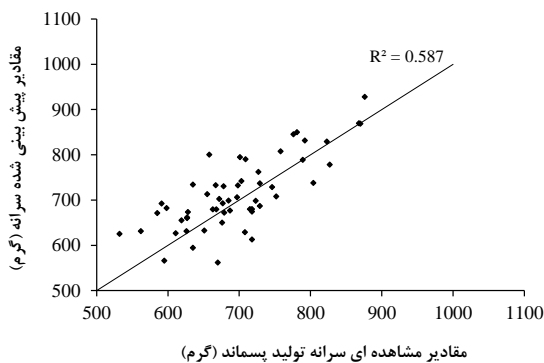
در این مقاله به منظور تدوین بهترین ساختار برای مدل M5P جهت پیش‌بینی سرانه تولید پسماند، از تکنیک اعتبارسنجی متقابل k- فولد (Cross validation) استفاده شده است. در این روش دسته داده‌ها به تعداد فولدهای مختلفی تقسیم شده و هر بار یک دسته به عنوان داده آزمایش و سایر دسته‌ها برای آموزش مدل به کار گرفته می‌شوند. جهت استخراج قواعد مربوطه در پیش‌بینی مقادیر سرانه تولید پسماند، تعداد فولدی که در روش صحت‌سنجی در مدل M5P، منجر به خطای پیش‌بینی کمتری گردد، مورد استفاده قرار می‌گیرد. در این مطالعه پس از ارزیابی مدل با تعداد فولدهای مختلف در نهایت مدل با تعداد ۱۳ فولد با مقدار خطای مجذور میانگین مربعات خطا (RMSE) برابر (gr/d) و ۵۵/۳۴ و میانگین قدر مطلق خطای نسبی (MARE) برابر ۶/۲۶ درصد



شکل ۴- ساختار درختی قواعد حاصل شده از مدل M5P در پیش‌بینی سرانه تولید پسماند

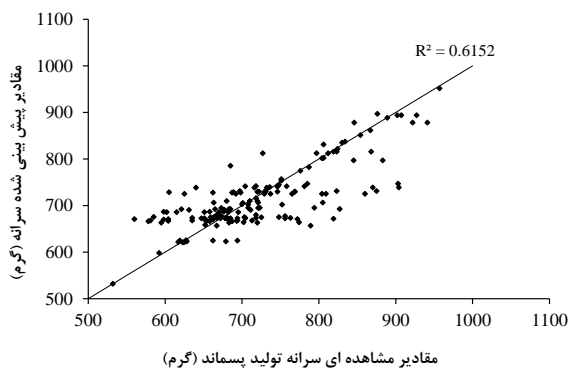
جدول ۱- قواعد حاصل از اجرای مدل M5P در پیش‌بینی سرانه تولید پسماند

LM num : 1
$SWG = 0.807 * P + 2.4379 * D - 0.1727 * E + 4.8887 * F + 562.2793$
LM num : 2
$SWG = 0.0516 * P + 2.4379 * D - 0.1727 * E + 4.8887 * F + 647.6783$
LM num : 3
$SWG = -0.2766 * P + 2.4379 * D - 0.1727 * E + 4.8887 * F + 763.2487$
LM num : 4
$SWG = -0.0001 * P + 2.6832 * D - 1.159 * E + 3.9024 * F + 740.0116$
LM num : 5
$SWG = -0.0002 * P + 4.6097 * D - 0.806 * E + 5.8302 * F + 797.9313$
LM num : 6
$SWG = -0.0003 * P + 14.9617 * D - 0.7183 * E + 5.8302 * F + 731.3821$
LM num : 7
$SWG = -0.0003 * P + 8.7545 * D - 0.7183 * E + 5.8302 * F + 738.8308$

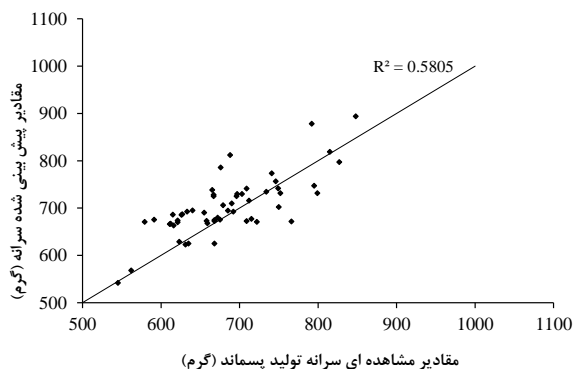


شکل ۷- مقایسه مقادیر مشاهده‌ای و پیش‌بینی شده نرخ تولید پسماندها توسط مدل MLP (داده‌های آزمایش)

با توجه به نمودارهای مقایسه مقادیر مشاهده‌ای و پیش‌بینی شده نرخ تولید پسماندها توسط مدل MLP مشخص می‌شود که این مدل‌ها در پیش‌بینی مقادیر سرانه با توجه به شاخص خطای MARE و RMSE عملکرد مناسبی را نشان داده‌اند. همچنین نتایج آموزش و صحت‌سنجی مدل SVM برای پیش‌بینی مقادیر سرانه تولید پسماند در شکل‌های (۸) و (۹) ارائه شده است. مقادیر سایر شاخص‌های آماری خطا برای ارزیابی عملکرد روش‌های بکار گرفته شده در جدول (۲) آورده شده است.

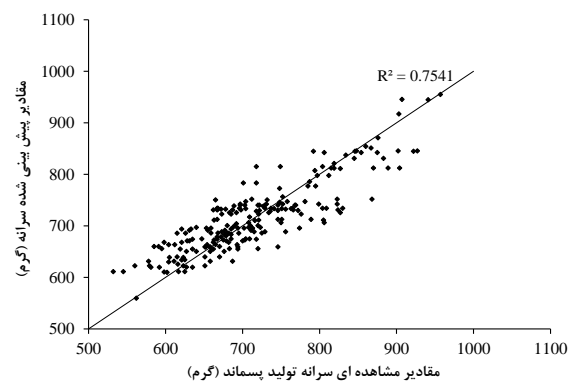


شکل ۸- مقایسه مقادیر مشاهده‌ای و پیش‌بینی شده نرخ تولید پسماندها توسط مدل SVM (داده‌های آموزش)



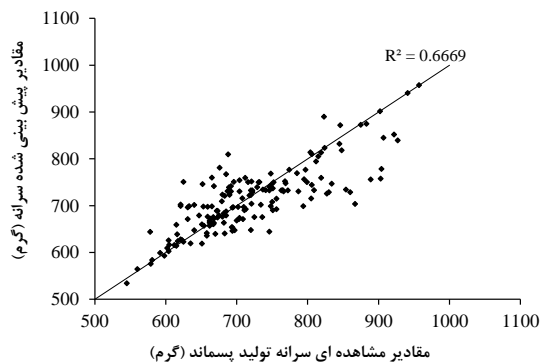
شکل ۹- مقایسه مقادیر مشاهده‌ای و پیش‌بینی شده نرخ تولید پسماندها توسط مدل SVM (داده‌های آزمایش)

همان طور که از نتایج مدل استنباط می‌شود، در وهله اول مدل در قیاس با سایر پارامترهای در نظر گرفته شده به پارامتر جمعیت (P) حساسیت بیشتری نشان داده است (شکل (۴)). با توجه به قواعد و قوانین حاصله از مدل M5P مشخص می‌باشد که مدل با ساختاری ساده و کارا به نحو مناسبی از طریق روابط خطی، موفق به پیش‌بینی سرانه تولید پسماند شده است که از مزیت‌های این روش نسبت به شبکه عصبی MLP می‌باشد. در شکل (۵)، مقادیر مشاهده‌ای و محاسباتی سرانه تولید پسماند توسط مدل M5P ارائه شده است.



شکل ۵- مقایسه مقادیر مشاهده‌ای و پیش‌بینی شده نرخ تولید پسماندها توسط مدل M5P

از مجموعه داده‌های اندازه‌گیری شده به ترتیب ۷۵ و ۲۵ درصد از داده‌ها به منظور آموزش و صحت‌سنجی مدل‌های MLP و SVM مورد استفاده قرار گرفته‌اند. خروجی مدل‌های MLP و SVM مقادیر نرخ تولید پسماندها را به دست می‌دهد. بدین ترتیب در ساختار مدل MLP به ازای تعداد لایه و نرون‌های متفاوت آنالیز حساسیت صورت گرفته است. مقادیر پارامترهای شبکه شامل تعداد نرون‌ها، بایاس‌ها و ... حین آموزش شبکه تنظیم شده‌اند. نتایج آموزش و صحت‌سنجی مدل MLP برای پیش‌بینی مقادیر سرانه تولید پسماند در شکل‌های (۶) و (۷) نشان داده شده است.



شکل ۶- مقایسه مقادیر مشاهده‌ای و پیش‌بینی شده نرخ تولید پسماندها توسط مدل MLP (داده‌های آموزش)

مورد استفاده قرار گیرد.

جدول ۲- خطای مدل‌های به کار گرفته شده در پیش‌بینی

سرانه تولید پسماند					
Models	MARE (%)	RMSE (gr/d)	CC	R ²	MAE (gr/d)
M5P	۶/۲۶۲	۵۵/۳۴۲	۰/۸۷	۰/۷۵۴	۴۴/۷۰۸
SVM	۷/۳۷	۶۶/۹۲	۰/۷۸	۰/۶۰	۵۲/۷۰
MLP	۸/۵۴۱	۷۰/۶۴۱	۰/۷۹	۰/۶۳۱	۵۸/۶۹۸

۶- قدردانی

بر خود لازم می‌دانیم از اعضاء محترم کارگروه مدیریت پسماند استان هرمزگان به ویژه معاونت عمرانی استانداری هرمزگان، دفتر امور شهری و شوراهای استانداری هرمزگان، دفتر امور روستایی و شوراهای استانداری هرمزگان و تمامی شهرداران و دهیاران استان هرمزگان، به جهت همکاری، کمال تشکر و قدردانی را به عمل آوریم.

۷- مراجع

- Abbasi M, Abdoli MA, Omidvar, B., Baghvand, A., "Forecasting Municipal Solid Waste Generation by Hybrid Support Vector Machine and Partial Least Square Model", International Journal of Environmental Research, winter 2013, Volume 7, Issue 1, 27-38.
- Abdoli MA, Falah Nezhad M, Salehi Sede R, Behboudian S, "Long term Forecasting of Solid Waste Generation by the Artificial Neural Networks", Environmental Progress & Sustainable Energy, December 2012, Volume 31, Issue 4, 628-636.
- Antanasijević D, Pocajt V, Popović I, Redžić N, Ristić M, "The forecasting of municipal waste generation using artificial neural networks and sustainability indicators", Sustainability Science, January 2013, Volume 8, Issue 1, 37-46.
- Avci E, "Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm-support vector machines: HGASVM", Expert Systems with Applications, March 2009, Volume 36, Issue 2, Part 1, 1391-1402.
- Bach H, Mild A, Natter M, Weber A, "Combining Socio-demographic and logistic factors to explain the generation and collection of waste paper", Resources, Conservation and Recycling, April 2004, Volume 41, Issue 1, 65-73.
- Batinić B, Vukmirović S, Vujić G, Stanisavljević N, Ubavin D, Vukmirović G, "Using ANN model to determine future waste characteristics in order to achieve specific waste management targets -case study of Serbia", Journal of Scientific and Industrial Research (JSIR), July 2011, 70(07), 513-518.
- Bayar S, Demir I, Engin GO, "Modeling leaching behavior of solidified wastes using back-propagation neural networks", Ecotoxicology and Environmental Safety, 2009, 72 (3), 843-850.
- Bogner J, Matthews E, "Global methane emissions from landfills: new methodology and annual estimates 1980-1996", Global Biogeochemical Cycles, June 2003, 17, NO. 2, 1-18.
- Chang N B, Lin Y T, "An analysis of recycling impacts on solid waste generation by time series intervention modeling", Resources, Conservation and Recycling, March 1997, Volume 19, Issue 3, 165-186.
- Chen HW, Chang N B, "Prediction analysis of solid waste generation based on grey fuzzy dynamic modeling", Resources, Conservation and Recycling, May 2000, Volume 29, Issues 1-2, 1-18.

با توجه به نتایج حاصل شده در این جدول مشخص می‌شود که روش M5P با مقدار مجذور میانگین مربعات خطا (RMSE) ۵۵/۳۴ (gr/d) و میانگین قدر مطلق خطای نسبی (MARE) ۶/۲۶ درصد، بهترین عملکرد را نسبت به سایر روش‌ها دارد. همچنین ضریب همبستگی بین نتایج در این حالت نسبت به روش‌های دیگر بیشتر شده است. بنابراین در قیاس با سایر مدل‌ها، مدل M5P می‌تواند با دقت مناسبی برای پیش‌بینی نرخ تولید پسماند به کار رود. نتایج پیش‌بینی داده‌های سرانه تولید پسماند در این روش رضایت بخش بوده و تطبیق مناسبی بین داده‌های واقعی و پیش‌بینی شده وجود دارد.

۵- بحث و نتیجه‌گیری

با توجه به این که کمیت پسماندهای یک شهر یا منطقه، لازمه برنامه‌ریزی در بخش‌های مختلف مدیریت پسماند است، آگاهی از مقدار آن بسیار حائز اهمیت است. به منظور طراحی اصولی سیستم مدیریت پسماند شهری و روستایی مناطق ساحلی جنوب ایران با افق ده ساله و به دلیل نبود آمار و اطلاعات از سرانه تولید پسماند، در این مطالعه ابتدا با هماهنگی انجام شده با مسئولان خدمات شهری و دهیاران محدوده مورد مطالعه، کل پسماند تولیدی تمامی شهرها و روستاها، به صورت هفت روز متوالی در وسط هر یک از چهار فصل سال ۱۳۹۲، اندازه‌گیری و پس از آنالیز و بررسی داده‌ها، سرانه تولید پسماند شهرها و روستاها به دست آمده است. سپس با در نظر گرفتن پارامترهای ارتفاع از سطح دریا، جمعیت، درجه شهری و تناوب جمع‌آوری پسماند، به عنوان داده‌های ورودی، مقادیر سرانه تولید پسماند با روش‌های SVM، MLP و M5P پیش‌بینی شده است. نتایج حاصله نشان می‌دهد که روش M5P نسبت به روش‌های قدیمی‌تر، شبکه عصبی و SVM کارایی بهتری دارد. از جمله ویژگی‌های بارز مدل‌های M5P ارائه روابط رگرسیونی و معادلات ریاضی کاربردی ساده و قابل فهم و در عین حال با دقت مناسب جهت تخمین و پیش‌بینی می‌باشد. این روش به دلیل سادگی و ارائه روابط ریاضی بین پارامترها و نیز دقت مناسب می‌تواند به عنوان ابزاری مفید و قدرتمند در پیش‌بینی سرانه تولید پسماند به ویژه در مناطقی که به دلیل بودجه ناکافی، تجهیزات اندازه‌گیری محدود و فقدان سیستم مدیریتی مناسب، با نبود داده‌های ثبت شده زمانمند از مقادیر پسماند مواجه هستند،

- Learning, 1989, 4(2), 227-243.
- Navarro-Esbrí J, Diamadopoulos E, Ginestar D, "Time series analysis and forecasting techniques for municipal solid waste management", *Resources, Conservation and Recycling*, May 2002, Volume 35, Issue 3, 201-214.
- Noori R, Abdoli MA, Jalili Ghazizade M, Samieifard R, "Comparison of neural network and principal component-regression analysis to predict the solid waste generation in Tehran", *Iranian Journal of Public Health*, 2009, Volume 38, Issue 1, 74-84.
- Noori R, Karbassi A, Sabahi MS, "Evaluation of PCA and Gamma test techniques on ANN operation for weekly solid waste prediction", *Journal of Environmental Management*, January-February 2010, Volume 91, Issue 3, 767-771.
- Quinlan JR, "Learning with continuous classes", in *Proceedings, AI'92, 5th Australian Joint Conference on Artificial Intelligence*, Adams & Sterling (eds.), World Scientific, Singapore, 1992, 343-348.
- Tchobanoglous G, Theisen H, Vigil SA, "Integrated Solid Waste Management: Engineering Principles and Management Issues", New York: McGraw-Hill Book Co., Inc, 1993.
- Vapnik V, "The Nature of Statistical Learning Theory (Information Science and Statistics)", New York, NY: Springer-Verlag, 1 edition, January 1995.
- Wang Y, Witten I, "Inducing model trees for continuous classes", In *Proceedings of the 9th European Conference on Machine Learning Poster Papers*, Prague, 1997, 128-137.
- Wanmg CM, Huang, Y. F., "Evolutionary- based feature selection approaches with new criteria for data mining: A case study of credit approval data", *Expert Systems with Applications*, April 2009, Volume 36, Issue 3, Part 2, 5900-5908.
- مینوسپهر م، "مطالعات طرح جامع مدیریت پسماند شهرها و روستاهای ساحلی استان هرمزگان"، مهندسین مشاور عمران محیط زیست شهر سبز پارسیان، استانداری هرمزگان، ۱۳۹۴-۱۳۹۱.
- Daskalopoulos E, Badr O, Probert SD, "Municipal solid waste: a prediction methodology for the generation rate and composition in the European Union countries and the United States of America", *Resources, Conservation and Recycling*, November 1998, Volume 24, Issue 2, 155-166.
- Dong C, Jin B, Li D, "Predicting the heating value of MSW with a feed forward neural network", *Waste Management*, 2003, Volume 23, Issue 2, 103-106.
- Dyson B, Chang NB, "Forecasting Municipal Solid Waste Generation in a Fast-Growing Urban Region with System Dynamics Modeling", *Waste Management*, 2005, Volume 25, Issue 7, 669-679.
- Hockett D, Lober DJ, Pilgrim K, "Determinants of per capita municipal solid waste generation in the Southeastern United States", *Journal of Environmental Management*, November 1995, Volume 45, Issue 3, 205-217.
- Jahandideh S, Asadabadi E, Askarian V, Movahedi MM, Hosseini S, Jahandideh M, "The use of artificial neural networks and multiple linear regression to predict rate of medical waste generation", *Waste Management*, November 2009, Volume 29, Issue 11, 2874-2879.
- Jenkins R R, "The Economics of Solid Waste Reduction: The Impact of User Fees (New Horizons in Environmental Economics)", Elgar, Edward Publishing, Inc, 1 edition, 1993.
- Kalogirou SA, "Artificial intelligence for the modeling and control of combustion processes: a review", *Progress in Energy and Combustion Science*, 2003, Volume 29, Issue 6, 515-566.
- Karaca F, Ozkaya B, "NN-LEAP: A neural network-based model for controlling leachate flow-rate in a municipal solid waste landfill site", *Environmental Modeling & Software*, August 2006, Volume 21, Issue 8, 1190-1197.
- Karavezyris V, Timpe KP, Marzi R, "Application of system dynamics and fuzzy logic to forecasting of municipal solid waste", *Mathematics and Computers in Simulation*, September 2002, Volume 60, Issues 3-5, 149-158.
- Mingers J, "An Empirical Comparison of Pruning Methods for Decision Tree Induction", *Machine*

EXTENDED ABSTRACTS

Performance Assessment of Computational Intelligence Techniques in Solid Waste Generation Forecasting (Case Study)

Mohammad Minousepehr, Mohammad Reza Alizadeh, Nasser Talebbeydokhti*

Department of Civil and Environmental Engineering, Shiraz University, Shiraz, Iran

Received: 04 August 2016; Accepted: 289 January 2017

Keywords:

Solid waste Generation forecasting, Multi-layer perceptron, Support vector machines, M5P model trees.

1. Introduction

Knowing the quantity of generated solid waste play a very significant role in solid waste management programs in a region. Due to lack of measured data as well as unavoidable errors in measurements, assessment of volume of generated solid waste is always challenging. Also, field measurement and continues monitoring of the volume of solid waste is usually costly, difficult and time-consuming. Accurate prediction of solid waste generation can be regarded as a key factor in future solid waste management system planning. Conventional forecasting methods in solid waste generation forecasting frequently use the demographic and socioeconomic factors in a per capita basis. In most cases, insufficient funds, the limited measuring equipment, lack of appropriate management systems and due to the lack of recorded data for the volume of generated solid waste cause many problems in integrated solid waste systems management (Dyson and Chang, 2005). In this study, three computational intelligence techniques including M5P model trees, support vector machines (SVM) and multi-layer perceptron (MLP) artificial neural network are utilized to predict solid waste generation in Hormozgan Province, Iran. After a sensitivity analysis, four more influential factors including elevation, population, urban development index (measures the level of development in cities based on infrastructure, the municipality established year, the metropolitan area, population, city product and income, health and education) and the frequency of garbage collection were considered in developing models. The performance of proposed models in solid waste generation forecasting are assessed via different error evaluation indices and finally the results are compared.

2. Methodology

2.1. Case study and data

The study area includes about 220 cities and Rural areas with an approximate length of 900 km in Hormozgan, a coastal province in south of Iran. This area begins with Koshkonar in the most western point of the area that is adjacent to the Persian Gulf and ends to the port of Jask, the easternmost point of the province which is adjacent to Oman Sea. In this study, due to the lack of recorded data for the volume of generated solid waste in the region, the per capita municipal solid waste generation was determined in different parts of the study area. By direct sampling of the generated waste in the major sources of waste production including residential and commercial areas in rural and urban areas during the four seasons of the years 2013-2014 (March 2013 to February 2014), a dataset for per capita solid waste generation was obtained.

* Corresponding Author

E-mail addresses: m_minousepehr@yahoo.com (Mohammad Minousepehr), alizadeh.mohamadreza@yahoo.com (Mohammad Reza Alizadeh), nassertaleb@gmail.com (Nasser Talebbeydokhti).

2.2. MLP

The neural models are basically based on the perceived process of the human brain. The artificial model of the brain is known as ANN. The theory and mathematical of Neural networks are mathematical models, which can assess the relation among inputs and outputs of a physical system (even if it is complex and nonlinear) through a network of attached nodes. ANN structure is called architecture, which is in a form arranged by neurons that are called layers. ANN customary architecture is composed of three layers: input layer, hidden layer and output layer. In our research, one hidden layer feedforward ANN was used for the SWG prediction. The Levenberg–Marquardt (L–M) algorithm was chosen for training the networks. In addition, feed-forward ANN was trained by the standard back propagation.

2.3. SVM

Kernel-based techniques such as support vector machines represent a major development in machine learning algorithms. Support vector machines (SVM) are a group of supervised learning methods that can be applied for classification or regression. Support vector machines represent an extension to nonlinear models of the generalized portrait algorithm developed by Vladimir Vapnik. The SVM algorithm is based on the statistical learning theory (Vapnik, 1995).

2.4. Model trees

M5P model trees were first introduced by Quinlan (1992) and then the idea was reconstructed and improved in a system called M5P by Wang and Witten (1997). An M5P model tree is an effective learning method for predicting real values. Model trees, like regression trees, are efficient for large data sets. However, model trees are generally much smaller and are more accurate than regression trees (Quinlan, 1992). The M5 model trees algorithm builds a regression tree by recursive splitting of the sample space to minimize the intra-subset variability in the values down from the root via branch to the node. The splitting processes continue up to the point that the output values fluctuate only slightly, or only a few samples remain. Remaining samples are output values of all samples that have reached the nodes. The variability is determined by standard deviation of the values that reach that node from the root through the branch. The standard deviation reduction (SDR) is calculated as follows:

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i) \quad (1)$$

Where T is the set of examples that reach the node, T_i the sets that are resulted from splitting the node according to the chosen attribute and sd is the standard deviation.

2.5. Evaluation of model accuracy

To evaluate the performance of developed models in forecasting values of per capita waste generation, four indices are assessed: root mean square error (RMSE) or (MAPE), mean absolute error (MAE), correlation coefficient (CC) and mean absolute relative error (MARE). The accuracy of the model is determined according to these criterias.

$$RMSE = \sqrt{\frac{1}{n} \sum (x_i - y_i)^2} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |(x_i - y_i)| \quad (3)$$

$$CC = \frac{\left(\frac{1}{n}\right) [(x_i - \mu_x)^T (y_i - \mu_y)]}{\sqrt{(1/n)(x_i - \mu_x)^2} \sqrt{(1/n)(y_i - \mu_y)^2}} \quad (4)$$

$$MARE = \frac{1}{n} \sum_{i=1}^n \left| \frac{(x_i - y_i)}{x_i} \right| \times 100 \quad (5)$$

In all formulas, the x_i represents the observation, the y_i represents the predicted value, n the number of observations, μ_x the mean of x and μ_y the mean of y .

3. Results and discussion

The Cross-validation method with a number of different folds was implemented to validate the M5P model and the model with 13-folds was selected to predict per capita solid waste generation in Hormozgan province due to its lowest prediction error indices. The M5P model with 13 Folds and the root mean square error (RMSE) of 55.34 (gr/d) and the mean absolute relative error (MARE) of 6.26 percent has reached the lowest error. In order to train and validate the SVM and MLP models, 75% and 25% of the dataset were used, respectively. Fig. 1 show the performance of M5P, MLP and SVM model in predicting the per capita waste generation. Also, Table 1. compares the different error evaluation indices of three models in solid waste generation forecasting. According to error indices, M5P model has a better performance in comparison to MLP and SVM.

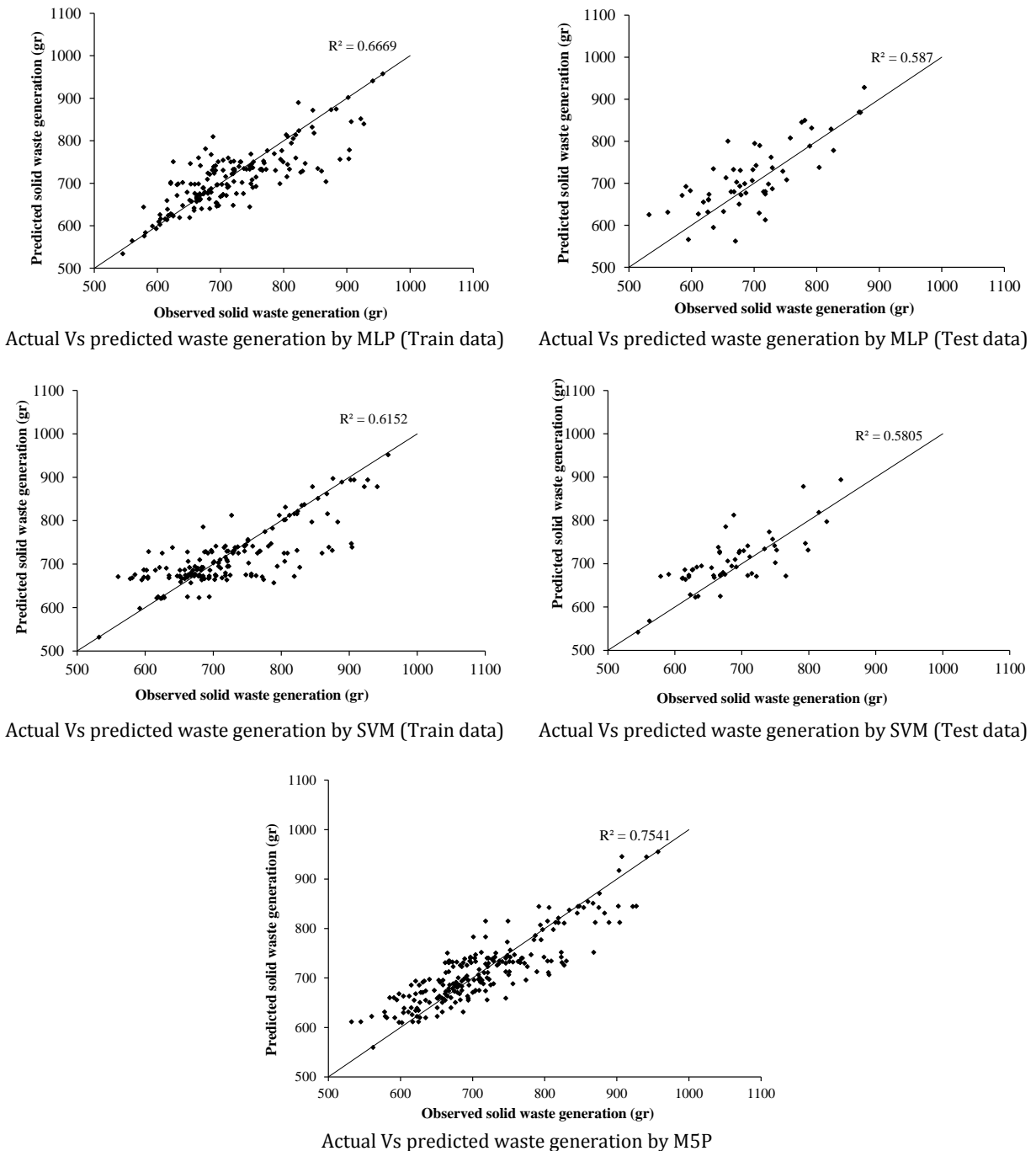


Fig. 1. Actual Vs predicted waste generation by different intelligence techniques

Table 1. Error of the used models to predict the per capita waste production

Models	MARE (%)	RMSE (gr/d)	CC	R ²	MAE (gr/d)
M5P	6.262	55.342	0.87	0.754	44.708
SVM	7.37	66.92	0.78	0.60	52.70
MLP	8.541	70.641	0.79	0.631	58.698

4. Conclusions

According to the results in Table 1, the M5P model with root mean square error (RMSE) (gr/d) 55.34 and the mean absolute relative error (MARE) 6.26 percent has the best performance in comparison with other methods. Furthermore, the correlation coefficient (CC) and R² between the results in this case is greater than other methods. Thanks to the simplicity and mathematical relationships between the parameters and the accuracy in predicting solid waste generation, the M5P model can be used as a useful and powerful tool. The comparison between predicted and observed date revealed that using intelligence techniques such as M5P model can be very practical in solid waste generation forecasting especially when there is not enough amount of recorded data for future integrated solid waste management and planning.

5. References

- Dyson B, Chang NB, "Forecasting Municipal Solid Waste Generation in a Fast-Growing Urban Region with System Dynamics Modeling", *Waste Management*, 2005, Volume 25, Issue 7, 669-679.
- Vapnik V, "The Nature of Statistical Learning Theory (Information Science and Statistics)", New York, NY: Springer-Verlag, 1 edition, January 1995.
- Quinlan JR, "Learning with continuous classes", in *Proceedings, AI'92, 5th Australian Joint Conference on Artificial. Intelligence*, Adams & Sterling (eds.), World Scientific, Singapore, 1992, 343-348.
- Wang Y, Witten I, "Inducing model trees for continuous classes", In *Proceedings of the 9th European Conference on Machine Learning Poster Papers*, Prague, 1997, 128-137.