

بسمه تعالی

گزارش درس اینترنت اشیا

استاد مربوطه : استاد محمد زارع

موضوع :

تمرین دوم درس اینترنت اشیا

نویسنده : میلاد سروری

Linear Regression:

R^2 : 0.998365740219586

MAE: 0.002456869950017464

MSE: 1.4792149428410072e-05

Decision Tree:

R^2 : 0.998619708765462

MAE: 0.0020226988827907006

MSE: 1.2493407988562736e-05

Neural Network:

R^2 : 0.998780392443753

MAE: 0.0021145682892146286

MSE: 1.1039014379619403e-05

تحلیل نتایج مدل‌ها:

با توجه به نتایج که شامل R^2 ، MAE و MSE برای سه مدل مختلف (رگرسیون خطی، درخت تصمیم و شبکه عصبی) است، در ادامه تحلیل دقیق‌تری از عملکرد این مدل‌ها و تأثیر ویژگی‌های مختلف بر پیش‌بینی مصرف انرژی انجام خواهیم داد.

نتایج مدل‌ها:

1. رگرسیون خطی: (Linear Regression)

$R^2 = 0.9984$ ○

MAE = 0.0025 ○

MSE = 1.479×10^{-5} ○

2. درخت تصمیم: (Decision Tree)

$R^2 = 0.9986$ ○

MAE = 0.0020 ○

MSE = 1.249×10^{-5} ○

3. شبکه عصبی: (Neural Network)

$R^2 = 0.9988$ ○

$$\begin{aligned} \text{MAE} &= 0.0021 \\ \text{MSE} &= 1.104 \times 10^{-5} \end{aligned}$$

کدام الگوریتم بهترین عملکرد را داشته است؟

با توجه به معیارهای مختلف، مدل شبکه عصبی بهترین عملکرد را در پیش‌بینی مصرف انرژی داشته است. دلیل آن را می‌توان در موارد زیر دید:

1. R^2 :

- شبکه عصبی با $R^2 = 0.9988$ بهترین توانایی را در توضیح واریانس داده‌ها داشته است. این یعنی 99.88% از واریانس داده‌های هدف توسط مدل شبکه عصبی توضیح داده شده است.
- درخت تصمیم نیز عملکرد خوبی داشته است ($R^2 = 0.9986$) و رگرسیون خطی با $R^2 = 0.9984$ کمی پایین‌تر از دیگر مدل‌ها است.

2. MAE (میانگین خطای مطلق):

- هر سه مدل خطای کمی دارند، اما درخت تصمیم با $\text{MAE} = 0.0020$ نسبت به دو مدل دیگر کمی بهتر عمل کرده است.
- شبکه عصبی و رگرسیون خطی به ترتیب با MAE های 0.0021 و 0.0025 در رتبه‌های بعدی قرار دارند.

3. MSE (میانگین مربعات خطا):

- شبکه عصبی با کمترین مقدار ($\text{MSE} = 1.104 \times 10^{-5}$) دقت بالاتری در پیش‌بینی‌ها داشته است. این مدل کمترین میزان خطا را در پیش‌بینی‌های خود نشان داده است.
- درخت تصمیم ($\text{MSE} = 1.249 \times 10^{-5}$) و رگرسیون خطی ($\text{MSE} = 1.479 \times 10^{-5}$) در رتبه‌های بعدی قرار دارند.

نتیجه‌گیری:

- با توجه به نتایج به دست آمده، مدل شبکه عصبی بهترین عملکرد را در تمام معیارها (R^2 ، MAE و MSE) داشته است و به نظر می‌رسد که بهترین گزینه برای پیش‌بینی مصرف انرژی در این مجموعه داده باشد. این مدل توانسته است بیشترین واریانس را توضیح دهد و کمترین خطای مربعی را در پیش‌بینی‌ها داشته باشد.

آیا ویژگی‌های خاصی بیشترین تأثیر را بر روی پیش‌بینی مصرف انرژی داشتند؟

برای تحلیل تأثیر ویژگی‌ها بر پیش‌بینی مصرف انرژی، می‌توان از روش‌های مختلف مانند اهمیت ویژگی‌ها در درخت تصمیم، ضریب‌های رگرسیون در رگرسیون خطی، یا ویژگی‌های یادگیری در شبکه‌های عصبی استفاده کرد. در ادامه، به تحلیل ویژگی‌های مختلف خواهیم پرداخت.

1. رگرسیون خطی:

در مدل رگرسیون خطی، تأثیر هر ویژگی بر پیش‌بینی مصرف انرژی از طریق ضریب‌های رگرسیون قابل شناسایی است. اگر ضریب ویژگی خاصی زیاد باشد، آن ویژگی تأثیر بیشتری بر پیش‌بینی خواهد داشت. به عنوان مثال:

- ویژگی‌هایی که ضریب بالاتری دارند، تأثیر بیشتری بر روی پیش‌بینی مصرف انرژی دارند.
- در صورتی که ضریب ویژگی‌ها منفی باشد، آن ویژگی موجب کاهش پیش‌بینی مصرف انرژی می‌شود.

2. درخت تصمیم:

درخت تصمیم می‌تواند با استفاده از اهمیت ویژگی‌ها نشان دهد که کدام ویژگی‌ها بیشتر در تقسیم‌بندی‌ها (split) استفاده شده‌اند. برای هر ویژگی، درخت تصمیم یک مقدار اهمیت (importance) محاسبه می‌کند که نشان‌دهنده میزان تأثیر آن ویژگی در پیش‌بینی خروجی است. به طور معمول، ویژگی‌هایی که بیشتر در گره‌های بالاتر درخت ظاهر می‌شوند، اهمیت بیشتری دارند.

در این مدل‌ها، ویژگی‌هایی مانند **Global_active_power**، **Global_reactive_power**، و **Global_intensity** معمولاً تأثیر زیادی بر پیش‌بینی مصرف انرژی دارند، چرا که این ویژگی‌ها به طور مستقیم با مصرف انرژی ارتباط دارند. همچنین، **Sub_metering_1**، **Sub_metering_2** و **Sub_metering_3** که به مصرف انرژی در بخش‌های مختلف مربوط می‌شوند، نیز می‌توانند تأثیر مهمی داشته باشند.

3. شبکه عصبی:

در شبکه‌های عصبی، تأثیر ویژگی‌ها به طور مستقیم از طریق وزن‌های مدل و عملکرد لایه‌های مخفی قابل شناسایی است. در این مدل‌ها، ویژگی‌هایی که در ابتدا (در لایه‌های اولیه) بیشتر فعال می‌شوند، ممکن است تأثیر بیشتری در پیش‌بینی‌ها داشته باشند.

برای درک بهتر تأثیر ویژگی‌ها در شبکه‌های عصبی، می‌توان از تکنیک‌هایی مانند تجزیه و تحلیل حساسیت (Sensitivity Analysis) یا نمودارهای SHAP (Shapley Additive Explanations) استفاده کرد که در آن‌ها اهمیت هر ویژگی به تفکیک مشخص می‌شود.

نتیجه‌گیری در مورد ویژگی‌ها:

- ویژگی‌هایی مانند **Global_active_power** و **Global_reactive_power** معمولاً بیشترین تأثیر را در پیش‌بینی مصرف انرژی دارند، زیرا به طور مستقیم مصرف انرژی را نشان می‌دهند.
- ویژگی‌های مرتبط با **Sub_metering_1**، **Sub_metering_2** و **Sub_metering_3** نیز می‌توانند تأثیر قابل توجهی در پیش‌بینی داشته باشند زیرا به مصرف انرژی در بخش‌های مختلف مرتبط هستند.

در نهایت، با استفاده از تحلیل‌های بیشتر مانند تحلیل حساسیت یا اهمیت ویژگی‌ها، می‌توان تأثیر دقیق‌تر هر ویژگی را بر پیش‌بینی مصرف انرژی شناسایی کرد.