

## عنوان مقاله:

اعمال مدل های رگرسیون بر زیرمجموعه های با همبستگی بالا برای بهبود جایگذاری مقادیر جافتاده عددی

## محل انتشار:

فصلنامه مهندسی برق دانشگاه تبریز، دوره 48، شماره 3 (سال: 1397)

تعداد صفحات اصل مقاله: 14

## نویسندگان:

امیرمسعود سفیدیان - دانشکده مهندسی کامپیوتر - دانشگاه تربیت دبیر شهید رجایی - تهران

نگین دانشپور - دانشکده مهندسی کامپیوتر - دانشگاه تربیت دبیر شهید رجایی - تهران

## خلاصه مقاله:

حضور مقادیر جافتاده در داده های دنیای واقعی مشکلی بسیار رایج و غیرقابل اجتناب است. بنابراین لازم است تا پیش از عملیات اکتشاف دانش، این مقادیر جافتاده به طور دقیق پر شوند. در این مقاله، سه رویکرد جدید برای تخمین مقادیر جافتاده عددی پیشنهاد می شود. در تمامی روش های پیشنهادی، مدل های رگرسیون بر زیرمجموعه هایی با همبستگی بالا اعمال می شوند. در انتخاب زیرمجموعه های مطلوب سعی می شود تا همبستگی بین صفت جافتاده و دیگر صفات حداکثر شود. انتخاب این زیرمجموعه ها با استفاده از رویکردهایی مبتنی بر انتخاب روبه جلو انجام می شود. از معیار ضریب همبستگی برای اندازه گیری میزان ارتباط بین صفات استفاده شده است. همچنین در روش های پیشنهادی، ترتیب صفات جافتاده برای انجام عمل جایگذاری اولویت دهی می شوند. عملکرد رویکردهای پیشنهادی بر روی پنج مجموعه داده از دنیای واقعی با مقادیر مختلف جافتادگی ارزیابی شده است. عملکرد رویکردهای ارائه شده با پنج رویکرد جایگذاری با مقدار میانگین، جایگذاری با استفاده از نزدیک ترین همسایگان، روش جایگذاری با خوشه بندی c-means فازی، روش جایگذاری با درخت تصمیم و روشی مبتنی بر رگرسیون به نام الگوریتم جایگذاری با رگرسیون افزایشی صفات (IARI) مقایسه شده است. از دو معیار شناخته شده ی ریشه میانگین مربعات خطا و ضریب تعیین برای مقایسه عملکرد رویکردهای پیشنهادی با دیگر روش های جایگذاری استفاده شده است. نتایج آزمایش ها نشان می دهد که رویکردهای ارائه شده، حتی زمانی که درصد جافتادگی بالا است، بهتر از دیگر روش های مقایسه شده عمل می کنند.

## کلمات کلیدی:

جایگذاری مقادیر جافتاده، همبستگی، رگرسیون

## لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/890121>

