



دانشگاه علوم پزشکی

و خدمات بهداشتی درمانی کرمان

دانشکده بهداشت

پایان نامه مقطع کارشناسی ارشد آمار زیستی

عنوان

مقایسه روش‌های K نزدیک‌ترین همسایگان، ماشین بردار پشتیبان و رگرسیون لجستیک در طبقه‌بندی افراد دارای بیرون زدگی دیسک کمر در استان کرمان در سال ۱۳۹۷

توسط

سمیرا جعفری خیرآبادی

استاد راهنما

خانم دکتر تانیا دهش

اساتید مشاور

خانم دکتر مینا جمشیدی / آقای دکتر فرهاد ایرانمنش

سال تحصیلی شهریور ۹۸

مقدمه و اهداف: سالیانه افراد زیادی با ناراحتی کمر درد به مراکز بهداشتی و درمانی مراجعه می‌کنند. این افراد در بسیاری از موارد برای تشخیص وجود بیرون زدگی دیسک کمر به مراکز تصویربرداری رزونانس مغناطیسی (MRI) ارجاع داده می‌شوند. انجام MRI غیر ضروری، بار مالی زیادی برای بیمار و سیستم بهداشتی به همراه دارد. هدف از انجام این مطالعه، طبقه‌بندی صحیح افراد با استفاده از سؤالات ساده دوحالتی (بله / خیر) در مورد شرایط بیمار و براساس سه روش کلاس‌بندی K نزدیک‌ترین همسایگان^۱، ماشین بردار پشتیبان^۲ و رگرسیون لجستیک^۳ به دو گروه دارای بیرون زدگی دیسک کمر و فاقد بیرون زدگی دیسک کمر می‌باشد.

مواد و روش‌ها: داده‌های این مطالعه در کلینیک بعثت، یکی از اصلی‌ترین کلینیک‌های درمانی در مرکز استان کرمان، در جنوب شرقی ایران جمع‌آوری شده‌اند که شامل ۴۰۰ نفر، ۲۵۳ نفر زن و ۱۴۷ نفر مرد می‌باشد که از کمر درد رنج می‌بردند. ویژگی‌های مورد بررسی شامل: سن، جنسیت، سابقه ضربه به کمر، داشتن اختلال حسی، داشتن اختلال حرکتی، سیگار کشیدن، سابقه مصرف مواد مخدر، وزن و قد افراد می‌باشد که با نظر فوق تخصص مغز و اعصاب گردآوری شده‌اند. پیامد مورد نظر ما، وجود بیرون زدگی دیسک کمر می‌باشد که یک پیامد دوحالتی است که نتیجه واقعی آن بعد از دیدن MRI بیماران، توسط پزشک فوق تخصص به صورت دارای بیرون زدگی / بدون بیرون زدگی مشخص می‌شود. همچنین برای بررسی اثر حجم نمونه، چهار حجم نمونه متفاوت ۵۰، ۱۰۰، ۱۵۰ و ۲۰۰ در نظر گرفته شده است تا بتوان ابتدا مناسب‌ترین مدل را با استفاده از داده‌های مجموعه آموزش به دست آورد و سپس کارایی این مدل را روی داده‌های آزمون بررسی نمود.

یافته‌ها: نتیجه برازش سه مدل آماری، نشان داد که در حجم نمونه ۵۰ و ۱۰۰ مدل ماشین بردار پشتیبان مدل مناسبی برای طبقه‌بندی افراد می‌باشد. همچنین در نمونه ۱۵۰ و ۲۰۰ تایی مدل K نزدیک‌ترین همسایگان، مناسب‌ترین مدل است زیرا هر پنج ملاک ارزیابی مربوط به این مدل بهترین مقدار را نسبت به دو مدل دیگر دارد. مدل رگرسیون لجستیک نیز در تمامی حجم‌های نمونه و با هر پنج ملاک ارزیابی نسبت به دو مدل دیگر از کارایی کمتری برخوردار است.

نتیجه‌گیری: با توجه به ملاک‌های ارزیابی یافته‌ها نشان دادند که مدل ماشین بردار پشتیبان در نمونه‌های ۵۰ و ۱۰۰ تایی و مدل K نزدیک‌ترین همسایگان در نمونه‌های ۱۵۰ و ۲۰۰ تایی بهترین کارایی را دارند. مدل رگرسیون لجستیک نیز ضعیف‌ترین مدل در این داده‌ها می‌باشد. به طور کلی نتیجه می‌گیریم که مدل ماشین بردار پشتیبان در حجم نمونه کم و با اطلاعات کم و مدل K نزدیک‌ترین همسایگان در حجم نمونه بالاتر، مدل‌های مناسبی برای طبقه‌بندی افراد می‌باشند. این روش‌های آماری می‌توانند کمک‌کننده مناسب

¹ K-Nearest Neighbors (KNN)

² Support Vector Machine (SVMs)

³ Logistic Regression (LR)

برای پزشکان در تشخیص بهتر بیماری باشند و از ارجاع غیر ضروری بیماران به مراکز MRI، ترس، نگرانی و بار مالی برای بیمار و همچنین تحمل هزینه‌های سنگین برای مراکز بهداشتی و درمانی اجتناب می‌شود.

واژه‌های کلیدی: طبقه‌بندی، K نزدیک‌ترین همسایگان، ماشین بردار پشتیبان، رگرسیون لجستیک، حساسیت، ویژگی، میزان شاخص طبقه‌بندی صحیح، بیرون زدگی دیسک کمر

Abstract

Introduction: Every year, people with low back pain (LBP) refer to health centers and are referred to MRI centers to diagnose lumbar disc herniation (LDH) in many cases. Doing unnecessary MRI involves a lot of financial burden for the patient and the health system. The purpose of this study is to classify the people correctly using simple questions (yes / no) about patient conditions, based on three classification methods, K- nearest neighbors (KNN), support vector machine (SVM) and logistic regression (LR) to two groups having lumbar disc herniation and not having lumbar disc herniation.

Methods: data was gathered at Besat clinic, one of the most important medical clinics in the center of Kerman province, southeast of Iran which included 252 women and 148 men suffering from LBP. The features of this study include: age, gender, waist injury, sensory disorders, movement disorder, smoking, history of drug use, weight and height which have been collected by the neurologist's opinion. The outcome variable is the presence of lumbar disc herniation, which is a binary variable and the result of it was being diagnosed after observing the result of MRI by the neurologist in the form of having lumbar disc/not having lumbar disc.

Results: The result of the fitting of three statistical models showed that support vector machine is a suitable model for classifying individuals in a small sample size (50 and 100).

Also, K-nearest neighbors is the most appropriate model as the sample size increases (150 and 200). Because all four criteria for evaluating this model have the best value compared to the other two models. Logistic regression model is also less efficient than the other two models in all sample sizes with all four evaluation criteria.

Conclusion: In the 50 and 100 samples, support vector machine, and K-nearest neighbors in the 150 and 200 samples have the best performance. Logistic regression model is also the weakest model in this data. In general, we conclude that support vector machine in the small sample size and the K-nearest neighbors in the higher sample size are suitable models for classifying individuals.



Kerman University of Medical Sciences

School of Public Health

In Prtial Fulfillment of the Requirements for the Degree Master of Science

Title:

**The Comparison of K-Nearest Neighbors, Support Vector Machine and
Logistic regression in classification of people with lumbar disc herniation in
Kerman province, 2018**

By:

Samira jafari

Supervisor:

Dr. Tania Dehesh

Advisers:

Dr. Mina Jamshidi

Dr. Farhad Iranmanesh

Year: 2019

تاریخ

بسمه تعالی

شماره

صورتجلسه دفاع از پایان نامه

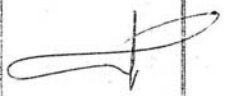
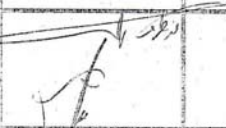
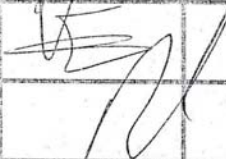


پیرست

استاد گروه بیوشیمی کرمان

استاد هیئت دانشکده

دفاعیه پایان نامه تحصیلی نامه خوانشمنند است نظر خود را در خانم سمیرا جمهری خیرابادی دانشجوی کارشناسی ارشد
بستی تحت عنوان "مقایسه روش های کال نزدیک ترین همسایگان ، ماشین بردار پشتیبان و رگرسیون لجستیک در طبقه
انفراد دارای بیرون زدگی دیسک کمر در استان کرمان در سال ۱۳۹۷" به راهنمایی خانم دکتر تانیا دهمش اعلام نمایم.

تاریخ ۱۱/۳۰ روز شنبه مورخ ۹۸/۶/۲۳ با حضور اعضای محترم هیات داوران متشکل از:

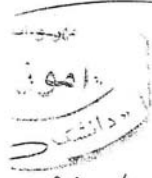
سمت	نام و نام خانوادگی	امضا
الف:استاد(ان) راهنما	خانم دکتر تانیا دهمش	
ب: استاد(ان) مشاور	۱-خانم دکتر مینا جمشیدی ۲-آقای دکتر فرهاد ایرانمنش	
ج: عضو هیات داوران (داخلی)	خانم دکتر مقدمه میرزایی	
د: عضو هیات داوران (خارجی)	آقای دکتر محمدرضا بانثی	
ه: نماینده تحصیلات تکمیلی	خانم دکتر مقدمه میرزایی	

بیت ۲۰۵

عالی

مورد تأیید

تاریخ و ضمن ارزیابی به شرح پیوست با درجه



۹۸-۶-۲۳