

## ارزیابی تاثیر کاهش داده در عملکرد طبقه بندی کننده های خطی و غیر خطی

حمید محمودآبادی<sup>۱</sup>، دکتر محمدرضا اصغری اسکویی<sup>۲</sup><sup>۱</sup> دانشگاه علامه طباطبایی، [Mahmoodabadihamid@atu.ac.ir](mailto:Mahmoodabadihamid@atu.ac.ir)<sup>۲</sup> دانشگاه علامه طباطبایی، [Oskoei@atu.ac.ir](mailto:Oskoei@atu.ac.ir)

## چکیده:

مهمی که در زمینه ارزیابی طبقه بندی کننده ها صورت پذیرفته اند، در زیر آورده شده اند.

لگون و همکارانش، پژوهشی انجام دادند که چند الگوریتم یادگیری (از جمله ماشین بردار پشتیبان) را روی مسائل تشخیص دستخط بر اساس دقت و هزینه محاسبه الگوریتم، مقایسه میکرد. [1] کوپر و همکارانش نتایجی از یک پژوهش ارائه دادند که عملکرد ۱۲ الگوریتم یادگیری را روی داده های پزشکی، با استفاده از معیارهای دقت و ROC میسنجید. [2]

لیم و همکارانش یک مقایسه ی تجربی بین الگوریتم درخت تصمیم و سایر الگوریتم های طبقه بندی انجام دادند که آنها را بر اساس معیار دقت میسنجید. [3]

هدف از ارائه این مقاله، بررسی میزان پایداری و عملکرد چهار الگوریتم طبقه بندی کننده بیزین، نزدیکترین همسایه، ماشین بردار پشتیبان و پرسپترون چند لایه بر روی مجموعه داده IRIS در برابر کاهش ابعاد داده ها توسط الگوریتم PCA میباشد. در این مقاله جهت بررسی تاثیر کاهش بعد روی نتایج الگوریتم ها، داده ها را در دو مرحله وارد الگوریتم نموده و دقت آنها را ثبت نمودیم، در مرحله اول، مجموعه داده با تعداد ابعاد پیش فرض را وارد الگوریتم نمودیم، در مرحله ی بعد، ابعاد داده ها را کاهش دادیم، آنرا وارد الگوریتم نموده و نتایج را ثبت نمودیم. در نهایت مقایسه ای بین نتیجه عملکرد الگوریتم در برابر داده ها با ابعاد پیش فرض و ابعاد کاهش یافته انجام دادیم.

در ادامه این مقاله، در بخش دوم، به توضیح مدلهای و روشهای به کار رفته خواهیم پرداخت. در بخش سوم به تشریح داده ها و آزمایشات انجام شده روی آنها و در بخش چهارم به جمع بندی و نتیجه گیری خواهیم پرداخت.

## روش ها و مدل ها

از پرکاربردترین ابزارها جهت خودکار سازی فرایندها، و شناسایی الگوها میتوان به یادگیری ماشین اشاره کرد. الگوریتم هایی نظیر بیزین، نزدیکترین همسایه، ماشین بردار پشتیبان، پرسپترون چند لایه و ... که با دقت بسیاری الگوها را از داده ها و اطلاعات استخراج کرده و به سیستم میدهند، تا سیستم با چگونگی برخورد با الگوهای ورودی

بررسی و مقایسه ی عملکرد الگوریتم های مختلف طبقه بندی در مسائل کاربردی یادگیری ماشین و شناخت الگو، همواره یک چالش بوده است. همچنین امروزه با افزایش میزان داده های تولید شده، در به کار گیری الگوریتم های طبقه بندی کننده، با مشکل بالا بودن حجم محاسبات مواجه هستیم. به همین علت، بکارگیری الگوریتم هایی که در پی کاهش ابعاد داده ها هستند، مورد توجه بسیاری قرار گرفته است. ما در این مقاله قصد داریم میزان پایداری و عملکرد چهار الگوریتم طبقه بندی کننده بیزین، نزدیکترین همسایه، ماشین بردار پشتیبان و پرسپترون چند لایه را بر روی مجموعه داده IRIS، در برابر کاهش ابعاد داده ها توسط الگوریتم PCA بررسی کنیم.

## واژه های کلیدی

ارزیابی طبقه بندی کننده ها، کاهش ابعاد، الگوریتم الگوریتم ماشین بردار پشتیبان، الگوریتم پرسپترون چند لایه، الگوریتم نزدیکترین همسایه

## مقدمه

در یادگیری ماشین و شناسایی الگو، طبقه بندی نوعی یادگیری با نظارت تلقی میشود که اشاره به نحوه حل مسئله تشخیص تعلق یک مشاهده جدید به یکی از کلاسها دارد. آموزش مدل با نظارت بر اساس مجموعه ای از داده هایی برچسب دار است که برچسب ها، تعلق مشاهدات به کلاس خاصی را تعیین میکند. بررسی و مقایسه ی عملکرد الگوریتم های مختلف طبقه بندی در مسائل کاربردی یادگیری ماشین و شناخت الگو، همواره یک چالش بوده است. همچنین امروزه با افزایش میزان داده های تولید شده، در به کار گیری الگوریتم های طبقه بندی کننده، با مشکل بالا بودن حجم محاسبات مواجه هستیم. به همین دلیل، کاربرد الگوریتم هایی که در پی کاهش ابعاد داده ها هستند، مورد توجه بسیاری قرار گرفته است. بطور خلاصه نتایج پژوهش های