

# گزارش فنی: مقایسه مدل‌ها (SVM و KNN) برای تشخیص اخبار جعلی

نویسنده: محمد امین کیانی 4003613052

## معرفی پروژه

در این پروژه، از دو مدل یادگیری ماشین، یعنی ماشین بردار پشتیبانی و نزدیک‌ترین همسایگی کدگذاری برای تشخیص اخبار جعلی استفاده شد. هدف از این پروژه ارزیابی عملکرد این دو مدل در زمینه تشخیص اخبار جعلی با استفاده از مجموعه داده‌های آموزشی و آزمون بود.

## مدل SVM

یک الگوریتم یادگیری ماشین است که برای دسته‌بندی داده‌ها استفاده می‌شود. در این مدل، هدف یافتن یک مرز هائپرپلین که داده‌ها را به دو گروه تقسیم کند. قادر به کار با داده‌های غیرخطی از طریق استفاده از هسته‌ها می‌باشد. در این پروژه از هسته استفاده شده است که به طور خاص در دسته‌بندی‌های پیچیده و غیرخطی که نیاز به عملکرد دقیق در فضاهای چندبعدی دارند، خوب عمل می‌کند.

## مدل KNN

یک مدل یادگیری نظارت شده است که برای طبقه‌بندی داده‌ها از فاصله بین داده‌ها استفاده می‌کند. این مدل ساده اما قدرتمند است و با توجه به تعداد کا نزدیک‌ترین همسایگان داده جدید، آن را به دسته‌بندی مناسب اختصاص می‌دهد. مدل عملکرد خوبی در مسائل ساده و داده‌های کمتر پیچیده دارد، اما ممکن است در داده‌های بزرگتر و پیچیده‌تر زمان اجرا بالا برود.

## مقایسه عملکرد مدل‌ها

در این پروژه، پس از اعمال پیش‌پردازش مناسب و تقسیم داده‌ها به مجموعه‌های آموزش و تست، مدل با هسته نسبت به مدل کا نزدیک‌ترین همسایه عملکرد بهتری داشت. دلیل این تفاوت در قدرت بالای آن در شناسایی الگوهای پیچیده در داده‌های غیرخطی است. مدل کا نزدیک‌ترین همسایه در داده‌های ساده‌تر عملکرد مناسبی داشت اما در مقایسه با در پردازش داده‌های پیچیده‌تر ضعیف‌تر بود.

## نتایج و ارزیابی

پس از ارزیابی مدل‌ها بر اساس معیارها مشخص شد که مدل بردار پشتیبانی در تمام این معیارها نسبت کا نزدیک‌ترین همسایه، عملکرد بهتری داشته است.

## ارزیابی دقیق‌تر مدل‌ها

مدل با هسته توانست با دقت بالاتر و تعداد بیشتری از نمونه‌ها را به درستی طبقه‌بندی کند با توجه به ویژگی‌های خاص هسته این مدل توانسته است مرزهای غیرخطی بین داده‌ها را بهتر شبیه‌سازی کند. از سوی دیگر، مدل کا نزدیک‌ترین همسایه با نیاز به محاسبات بیشتر برای تعیین نزدیک‌ترین همسایگان، در هنگام داده‌های پیچیده‌تر به دقت پایین‌تری رسید.

## مقایسه در نمودارها

در این بخش، نمودارهای دقت و لاس برای هر دو مدل به نمایش گذاشته شده‌اند تا مقایسه‌ی عملکرد این دو مدل به صورت بصری قابل مشاهده باشد. با توجه به نتایج، مدل بردار پشتیبان در تمام مراحل عملکرد بهتری از خود نشان داد.

## نتایج مدل‌ها در جدول

مدل	Accuracy	Precision	Recall	F1 Score
KNN	0.89	0.90	0.88	0.89
SVM	0.99	0.99	0.99	0.99

  

KNN Classification Report:					
	precision	recall	f1-score	support	
0	0.89	0.91	0.90	3501	
1	0.90	0.88	0.89	3234	
accuracy			0.90	6735	
macro avg	0.90	0.90	0.90	6735	
weighted avg	0.90	0.90	0.90	6735	

  

SVM Classification Report:					
	precision	recall	f1-score	support	
0	1.00	0.99	1.00	3501	
1	0.99	1.00	0.99	3234	
accuracy			0.99	6735	
macro avg	0.99	1.00	0.99	6735	
weighted avg	0.99	0.99	0.99	6735	

