نشریه پژوهش در نشخوارکنندگان
جلد چهارم، شماره چهارم، ۱۳۹۵
http://ejrr.gau.ac.ir

تعیین مهم‌ترین عوامل مؤثر بر چربی شیر گاوهای هلشتاین با استفاده از الگوریتم انتخاب ویژگی

شهریار کارگر[1] و مرضیه مکرم[2]

[1]استادیار گروه علوم دامی، دانشکده کشاورزی، دانشگاه شیراز،

[2]استادیار گروه مرتع و آبخیزداری، دانشکده کشاورزی و منابع طبیعی داراب، دانشگاه شیراز

تاریخ دریافت: ۹۵/۸/۱۳؛ تاریخ پذیرش: ۹۵/۱۰/۲۹

چکیده

سابقه و هدف: پایش ترکیبات شیر یک گله می‌تواند به ارزیابی سلامت و وضعیت تغذیه‌ای گاوهای شیری کمک کند. چربی شیر یکی از با ارزش‌ترین ترکیبات شیر به شمار می‌رود. چربی متغیرترین ترکیب شیر بوده و عوامل بسیار متعدد فیزیولوژیکی و محیطی می‌توانند نسبت و مقدار آن را در شیر تحت تأثیر قرار بدهند. در صورتی که چربی شیر گله‌ای ۰/۳ درصد کم‌تر از میانگین نژادی باشد می‌تواند نشان‌گر بروز مشکلی باشد. جیره‌هایی که امکان بهبود در خروج بیش‌تر چربی را در شیر فراهم نمایند می‌توانند به‌طور بالقوه از نظر اقتصادی سودمند باشند. بنابراین، جهت حفظ چربی شیر ضروری است که مهم‌ترین عوامل مؤثر بر آن شناسایی، پایش و اندازه‌گیری شود. الگوریتم انتخاب ویژگی یکی از روش‌های داده‌کاوی است و به زعم دانسته ما در هیچ مقاله‌ای از آن برای واکاوی مهم‌ترین عوامل مؤثر بر چربی شیر گاوهای هلشتاین استفاده نشده است. از این‌رو، هدف از این پژوهش تعیین مهم‌ترین عوامل مؤثر بر چربی شیر گاوهای هلشتاین با استفاده از الگوریتم انتخاب ویژگی بود.

مواد و روش‌ها: برای انجام این پژوهش از ۲۱۱۲ داده خام حاصل از ۶۶ عامل بالقوه مؤثر بر چربی شیر استفاده شد. داده با سه روش مهم الگوریتم انتخاب ویژگی (روش‌های Best-First، Greedy-Stepwise و Ranker) و شش مدل مختلف (CFS-Subset-Eval، Info-Gain-Attribute-Eval، Gain-Ratio-Attribute-Eval،

*مسئول مکاتبه: skargar@shirazu.ac.ir

Symmetricer-Attribute-Eval، ReliefF-Attribute-Eval و Principal-Components) جهت تعیین مهم‌ترین عوامل مؤثر بر چربی شیر ارزیابی شد. داده با استفاده از نرم‌افزار WEKA (نسخه ۳/۸) واکاوی شد.

یافته‌ها: نتایج نشان داد که روش‌های Best-First و Greedy-Stepwise الگوریتم انتخاب ویژگی مناسب‌ترین روش برای انتخاب عوامل مؤثر بر چربی شیر با استفاده از طبقه‌بندی Naive Bayes با کم‌ترین میزان خطا هستند. بر این اساس، مهم‌ترین عوامل مؤثر بر چربی شیر به ترتیب اسیدهای چرب غیر استریفیه خون، کل اسیدهای چرب فرار مایع شکمبه، عصاره اتری مصرفی، کل زمان صرف شده برای فعالیت نشخوار به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، کل زمان صرف شده برای فعالیت نشخوار به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی علوفه‌ای مصرفی، کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، وزن بدن، نمره وضعیت بدنی، ضخامت لایه چربی پشت و گلوکز خون تعیین شدند.

نتیجه‌گیری: بر اساس انتخاب تعداد محدودی از عوامل مؤثر بر چربی شیر با الگوریتم انتخاب ویژگی امکان پایش و اندازه‌گیری آن‌ها جهت حفظ مؤثر چربی شیر در مزرعه و صرفه‌جویی در زمان و هزینه وجود دارد.

واژه‌های کلیدی: چربی شیر، الگوریتم انتخاب ویژگی، گاو شیری

مقدمه

چربی متغیرترین ترکیب شیر بوده و عوامل بسیار متعدد فیزیولوژیکی و محیطی آن را تحت تأثیر قرار می‌دهند. جیره‌هایی که امکان بهبود در خروج بیش‌تر چربی را در شیر فراهم نمایند می‌توانند به طور بالقوه از نظر اقتصادی سودمند باشند. از این‌رو، جهت حفظ چربی شیر ضروری است که مؤثرترین عوامل مؤثر بر چربی شیر شناسایی، پایش و اندازه‌گیری شود. برای تعیین مهم‌ترین عوامل مؤثر بر یک مسأله خاص، راه حل‌ها و الگوریتم‌های فراوانی می‌تواند وجود داشته باشد که مشکل بعضی از این الگوریتم‌ها بار محاسباتی زیاد آن‌ها است. اگر چه امروزه با ظهور کامپیوترهای نسل جدید با قدرت پردازش سریع و منابع ذخیره‌سازی بزرگ این مشکل تا حدودی مرتفع شده است اما وجود مجموعه داده‌ای بسیار بزرگ برای مسائل جدید (که گاه اندازه‌گیری آن‌ها ناممکن و یا سخت و بعضاً زمان‌بر و پر هزینه است) سبب شده است که هم‌چنان پیدا کردن یک الگوریتم دقیق و سریع‌ای که بتواند با معرفی تعداد محدودی از عوامل تأثیرگذار به‌عنوان خروجی بیش‌ترین تغییرات مربوط به یک متغیر خاص را توجیه نماید، مهم باشد. از این رو، هدف از این پژوهش تعیین مهم‌ترین عوامل مؤثر بر چربی شیر گاوهای هلشتاین با استفاده از الگوریتم انتخاب ویژگی (Feature Selection) بود.

مواد و روش‌ها

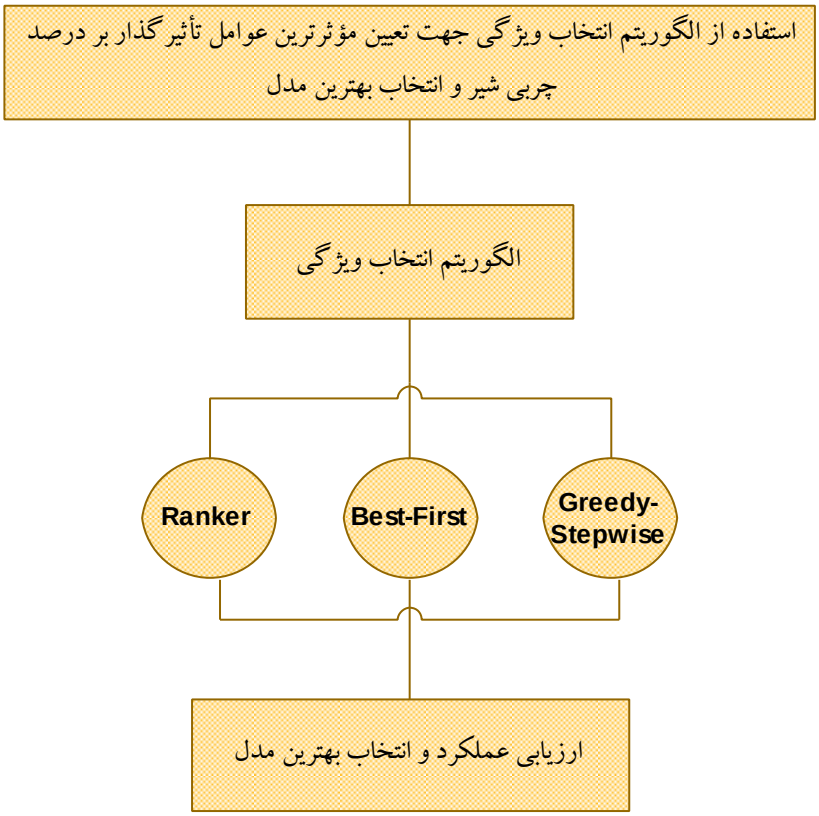
برای انجام این پژوهش از ۲۱۱۲ داده خام حاصل از ۶۶ عامل بالقوه مؤثر بر چربی شیر استفاده شد (۵، ۶، ۷ و ۸). داده با سه روش مهم الگوریتم انتخاب ویژگی (روش‌های Best-First، Greedy-Stepwise و Ranker) جهت تعیین مهم‌ترین عوامل مؤثر بر چربی شیر با استفاده از نرم‌افزار WEKA (نسخه ۳/۸) واکاوی شد (۳). در نگاره ۱ روند عمل الگوریتم انتخاب ویژگی نمایش داده شده است. روش‌های مختلف انتخاب ویژگی تلاش می‌کنند تا زیر مجموعه‌ای با کمینه تعداد ویژگی (عوامل تأثیرگذار) ممکن برای پیش‌بینی هدف موردنظر پیدا کنند. در تمام این روش‌ها بر اساس کاربرد و نوع تعریف، زیر مجموعه‌ای به‌عنوان جواب انتخاب می‌شود که بتواند مقدار یک تابع ارزیابی را بهینه کند. فرآیند انتخاب ویژگی در تمامی روش‌ها به بخش‌های پی‌آیند تقسیم می‌شود:

۱- تابع تولید کننده: این تابع زیر مجموعه‌های کاندید را برای روش موردنظر پیدا می‌کند.

۲- تابع ارزیابی: زیرمجموعه موردنظر را بر اساس روش داده شده، ارزیابی و یک عدد به‌عنوان میزان خوبی روش باز می‌گرداند. روش‌های مختلف سعی در یافتن زیرمجموعه‌ای دارند که این مقدار را بهینه کند.

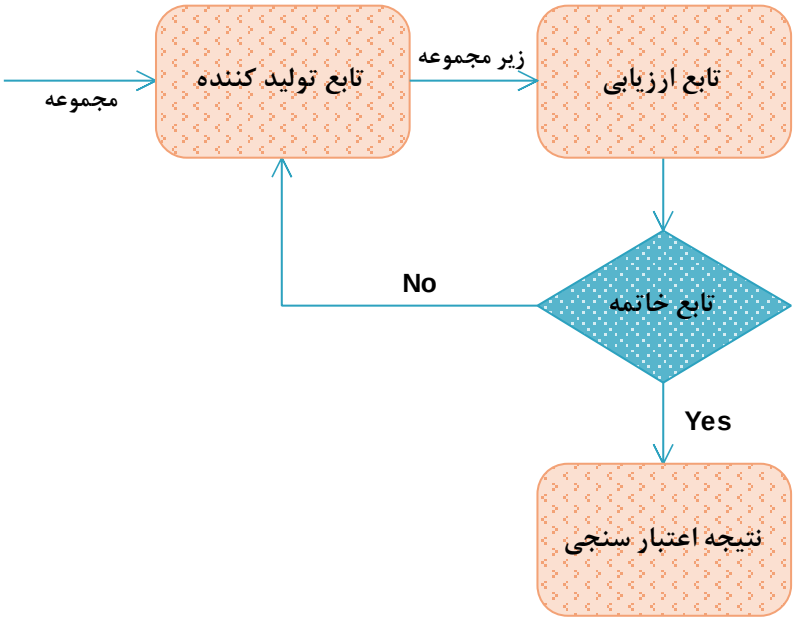
۳- شرط خاتمه: برای تصمیم‌گیری در مورد زمان توقف الگوریتم.

در ادامه به کمک تابع تعیین اعتبار مشخص می‌شود که آیا زیر مجموعه انتخاب شده معتبر است یا خیر. گامه تعیین اعتبار به کمک این تابع در شکل ۲ نشان داده شده است.



شکل ۱- روند الگوریتم انتخاب ویژگی به‌منظور انتخاب مؤثرترین عوامل تأثیرگذار بر درصد چربی شیر.

Figure 1. Feature selection algorithm process to select the most efficient factors affecting milk fat percentage.



شکل ۲- گامه‌های الگوریتم انتخاب ویژگی.

Figure 2. Feature selection algorithm steps.

درخت‌های تصمیم روشی برای نمایش یک سری از قوانین هستند که منتهی به یک رده یا مقدار می‌شوند. درخت‌های تصمیمی که برای پیش‌بینی متغیرهای دسته‌ای استفاده می‌شوند، درخت‌های طبقه‌بندی (Classification) نامیده می‌شوند. الگوریتم‌های درخت تصمیم که در این پژوهش مورد استفاده قرار گرفته‌اند عبارتند از:

الف- الگوریتم J48: این الگوریتم سعی می‌کند تا به صورت بازگشتی مجموعه داده را با استفاده از اطلاعات نرمال به زیر مجموعه‌هایی تقسیم کند که از ریشه به سمت برگ حرکت کرده و در نهایت به طبقه موردنظر می‌رسد. تولید درخت تصمیم اولیه از مجموعه داده مهم‌ترین بخش این الگوریتم محسوب می‌شود. این الگوریتم در نهایت یک طبقه‌بندی انجام داده که دارای دو گره می‌باشد. یک گره به‌صورت برگ که یک دسته را مشخص می‌کند و یک گره تصمیم که آزمون‌هایی روی یک صفت انجام می‌دهد تا یک شاخه به ازای هر خروجی آزمون تولید کند.

ب- الگوریتم Naive Bayes: این الگوریتم برای طبقه‌بندی داده از احتمال وقوع یا عدم وقوع پدیده‌ها استفاده می‌کند. در واقع، براساس ویژگی‌های ذاتی احتمال با آموزش و ارزیابی داده در نهایت داده را به زیر مجموعه‌هایی طبقه‌بندی می‌کند.

ج- الگوریتم LMT: این الگوریتم یک مدل طبقه‌بندی است که ترکیبی از پیش‌بینی لجستیک و یادگیری درخت تصمیم‌گیری لجستیک (LMT[1]) است (۱۰). این الگوریتم بر اساس رابطه متغیرهای وابسته و مستقل و در نظر گرفتن میزان خطا، داده را به زیر مجموعه‌هایی طبقه‌بندی می‌کند.

روش‌های مختلفی به‌منظور بررسی روش‌های انتخاب داده و انتخاب بهترین روش وجود دارد که از معروف‌ترین این روش‌ها می‌توان به AMS[2] و ARAE[3] اشاره کرد که در این مطالعه استفاده شده است (۱۱). در روش AMS به‌منظور بررسی دقت مدل از معادله زیر استفاده می‌شود. این روش نسبت بین میانگین تعداد کلاس‌های نادرست پیش‌بینی شده توسط مدل به تعداد کل داده است (۱).

$$AMS = \frac{\sum_{i=1}^{n} MS_i}{N} \qquad (۱)$$

که MS_i تعداد کلاس‌های نادرست پیش‌بینی شده توسط مدل و N تعداد کل داده است. در روش ARAE از خطای مطلق نسبی به کل داده استفاده می‌شود (۱).

$$ARAE = \frac{\sum_{i=1}^{n} RAE_i}{N} \qquad (۲)$$

که RAE_i خطای مطلق نسبی و N تعداد کل داده است. مجموع کل داده مورد مطالعه در این پژوهش و ویژگی هر کدام از آن‌ها (کمینه، بیشینه، میانگین و انحراف معیار) در جدول ۱ آورده شده است.

1- Logistic model tree
2- Average number of misclassified samples
3- Average relative absolute error

جدول ۱- ویژگی داده مورد مطالعه.

Table 1. Characteristics of the study data.

ویژگی Characteristic		کمینه Minimum	بیشینه Maximum	انحراف معیار SD	میانگین Average	ویژگی Characteristic		کمینه Minimum	بیشینه Maximum	انحراف معیار SD	میانگین Average
کل زمان صرف شده برای فعالیت خوردن به ازای هر وعده غذایی (دقیقه)	Eating time (min/bouts)	8.4	26.0	4.3	15.2	نسبت استات به پروپیونات	Acetate:propionate	1.90	4.35	0.63	3.03
کل زمان صرف شده برای فعالیت خوردن به ازای هر کیلوگرم ماده خشک مصرفی (دقیقه)	Eating time (min/DMI)	8.0	14.5	1.7	11.1	استات (میلی‌مولار)	Acetate (m*M*)	51.26	78.20	6.49	66.24
کل زمان صرف شده برای فعالیت خوردن به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی (دقیقه)	Eating time (min/NDFI)	27.3	46.6	5.5	36.2	پروپیونات (میلی‌مولار)	Propionate (m*M*)	16.58	32.99	4.51	22.68
کل زمان صرف شده برای فعالیت خوردن به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی علوفه‌ای مصرفی (دقیقه)	Eating time (min/fNDFI)	40.2	72.7	8.2	55.7	بوتیرات (میلی‌مولار)	Butyrate (m*M*)	7.07	22.38	3.16	12.39
کل زمان صرف شده برای فعالیت خوردن به ازای هر کیلوگرم NDF مؤثر فیزیکی مصرفی برای ذرات درشت‌تر از ۸ میلی‌متر (دقیقه)	Eating time (min/peNDF$_{>8}$ intake)	77.2	161.5	18.9	111.1	کل اسیدهای چرب فرار (میلی‌مولار)	Total volatile fatty acid (m*M*)	85.46	126.84	10.58	101.33
کل زمان صرف شده برای فعالیت خوردن به ازای هر کیلوگرم NDF مؤثر فیزیکی مصرفی برای ذرات درشت‌تر از ۱/۱۸ میلی‌متر (دقیقه)	Eating time (min/peNDF$_{>1.18}$ intake)	35.3	62.6	7.6	47.6	pH مایع شکمبه	Rumen fluid pH	5.98	6.66	0.20	6.33
کل فعالیت جویدن (دقیقه در روز)	Total chewing activity (TCA; min/d)	561.0	831.0	72.3	680.4	ماده خشک مصرفی (کیلوگرم در روز)	Dry matter intake (kg/d)	18.00	29.21	2.56	22.82
کل زمان صرف شده برای فعالیت جویدن به ازای هر وعده غذایی (دقیقه)	TCA (min/bouts)	26.0	79.6	12.6	41.2	ماده آلی مصرفی (کیلوگرم در روز)	Organic matter intake (kg/d)	16.45	26.46	2.31	20.73
کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم ماده خشک مصرفی (دقیقه)	TCA (min/DMI)	24.6	40.2	3.6	30.0	پروتئین خام مصرفی (کیلوگرم در روز)	Crude protein intake (kg/d)	2.70	4.59	0.46	3.65
کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی (دقیقه)	TCA (min/NDFI)	76.3	128.6	11.4	97.5	الیاف نامحلول در شوینده خنثی مصرفی (کیلوگرم در روز)	Neutral detergent fiber intake (kg/d)	5.34	9.43	0.87	7.04

ادامه جدول ۱

کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی علوفه‌ای مصرفی (دقیقه)	TCA (min/fNDFI)	123.1	200.8	150.2	17.9	الیاف نامحلول در شوینده اسیدی مصرفی (کیلوگرم در روز)	Acid detergent fiber intake (kg/d)	2.98	5.61	0.67	4.01
کل زمان صرف شده برای فعالیت جویدن به ازای مؤثر فیزیکی مصرفی برای NDF هر کیلوگرم ذرات درشت‌تر از ۸ میلی‌متر (دقیقه)	TCA (min/peNDF$_{>8}$ intake)	220.0	446.0	300.7	50.1	عصاره اتری مصرفی (کیلوگرم در روز)	Ether-extract intake (kg/d)	0.46	1.32	0.17	0.75
کل زمان صرف شده برای فعالیت جویدن به ازای مؤثر فیزیکی مصرفی برای NDF هر کیلوگرم ذرات درشت‌تر از ۱/۱۸ میلی‌متر (دقیقه)	TCA (min/peNDF$_{>1.18}$ intake)	98.9	160.5	128.2	15.1	کربوهیدرات‌های غیر الیافی مصرفی (کیلوگرم در روز)	Non-fibrous carbohydrate intake (kg/d)	7.34	11.13	0.88	9.20
زمان صرف شده برای استراحت (دقیقه در روز)	Resting time (min/d)	609.0	879.0	759.6	72.3	الیاف نامحلول در شوینده خنثی علوفه‌ای مصرفی (کیلوگرم در روز)	Forage NDF intake (kg/d)	3.60	5.84	0.51	4.56
زمان صرف شده برای فعالیت نشخواری (دقیقه در روز)	Ruminating time (min/d)	332.0	536.0	428.7	52.7	ذرات باقیمانده خوراک بر روی الک ۱۹ میلی‌متر (درصد)	Particles retained on 19 mm sieve (%)	3.53	11.02	1.87	7.80
زمان صرف شده برای فعالیت خوردن (دقیقه در روز)	Eating time (min/d)	185.0	301.0	251.8	31.4	ذرات باقیمانده خوراک بر روی الک ۸ میلی‌متر (درصد)	Particles retained on 8 mm sieve (%)	19.89	33.88	3.96	25.88
وزن بدن (کیلوگرم)	Body weight (kg)	612.2	772.4	671.4	48.2	ذرات باقیمانده خوراک بر روی الک ۱/۱۸ میلی‌متر (درصد)	Particles retained on 1.18 mm sieve (%)	33.00	50.24	6.02	41.35
نمره وضعیت بدنی	Body condition score	1.50	3.50	2.60	0.55	ذرات باقیمانده خوراک بر روی سینی (درصد)	Particles retained on Pan (%)	21.83	29.16	2.30	24.97
ضخامت لایه چربی پشت (میلی‌متر)	Back fat thickness (mm)	17.0	30.0	23.5	2.8	میانگین هندسی اندازه ذرات خوراک (میلی‌متر)	Geometric mean particle size (mm)	3.73	4.66	0.22	4.09
انسولین (میکرو واحد بین المللی بر میلی‌متر)	Insulin (µIU/mL)	7.26	26.61	11.82	4.57	تعداد وعده‌های فعالیت نشخواری (تعداد در روز)	Rumination (Bout/d)	10.0	18.0	1.8	14.0
گلوکز (میلی‌گرم بر دسی‌لیتر)	Glucose (mg/dL)	29.0	61.0	45.18	8.53	زمان صرف شده برای هر وعده فعالیت نشخواری (دقیقه)	Rumination length (min)	23.0	42.0	4.1	31.0

ادامه جدول ۱

69.6	13.8	106.0	51.0	Rumination interval (min)	فاصله زمانی بین وعده‌های فعالیت نشخواری (دقیقه)	0.166	0.065	0.412	0.046	Non-esterified fatty acid (mEq/L)	اسیدهای چرب غیر استریفیه (میلی‌اکی‌والان بر لیتر)
31.0	4.0	41.8	23.4	Ruminating time (min/bouts)	کل زمان صرف شده برای فعالیت نشخواری به ازای هر وعده (دقیقه)	139.6	42.3	206.0	62.0	Cholesterol (mg/dL)	کلسترول (میلی‌گرم بر دسی‌لیتر)
18.9	2.43	26.2	15.0	Ruminating time (min/DMI)	کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم ماده خشک مصرفی (دقیقه)	10.6	2.1	16.0	6.0	Triglyceride (mg/dL)	تری‌گلیسرید (میلی‌گرم بر دسی‌لیتر)
61.3	7.5	82.4	46.4	Ruminating time (min/NDFI)	کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی (دقیقه)	0.27	0.07	0.44	0.14	Total antioxidant capacity (mmol/L)	ظرفیت آنتی‌اکسیدانی کل (میلی‌مول بر لیتر)
94.5	12.1	130.9	74.8	Ruminating time (min/fNDFI)	کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی علوفه‌ای مصرفی (دقیقه)	1.723	0.644	3.125	0.875	Malondialdehyde (nmol/mL)	مالون دی آلدهاید (نانومول بر میلی‌لیتر)
189.6	34.8	284.4	135.6	Ruminating time (min/peNDF$_{>8}$ intake)	کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم NDF مؤثر فیزیکی مصرفی برای ذرات درشت‌تر از ۸ میلی‌متر (دقیقه)	2.30	0.36	3.42	1.86	peNDF$_{>8}$ intake (kg/d)	مصرف NDF مؤثر فیزیکی برای ذرات درشت‌تر از ۸ میلی‌متر (کیلوگرم در روز)
80.6	9.6	104.6	60.1	Ruminating (min/peNDF$_{>1.18}$ intake)	کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم NDF مؤثر فیزیکی مصرفی برای ذرات درشت‌تر از ۱/۱۸ میلی‌متر (دقیقه)	5.37	0.74	7.27	3.99	peNDF$_{>1.18}$ intake (kg/d)	مصرف NDF مؤثر فیزیکی برای ذرات درشت‌تر از ۱/۱۸ میلی‌متر (کیلوگرم در روز)
17.6	4.0	29.0	10.0	Meals (Bout/d)	تعداد وعده‌های غذایی (تعداد در روز)	1.76	0.52	3.04	0.65	Intake of particles retained on 19 mm sieve (kg/d)	مصرف ذرات باقیمانده بر روی الک ۱۹ میلی‌متر (کیلوگرم در روز)
15.1	4.3	26.0	8.0	Meal length (min)	زمان صرف شده برای هر وعده غذایی (دقیقه)	5.79	1.18	8.59	4.12	Intake of particles retained on 8 mm sieve (kg/d)	مصرف ذرات باقیمانده بر روی الک ۸ میلی‌متر (کیلوگرم در روز)
74.8	19.8	125.0	40.0	Meal interval (min)	فاصله زمانی بین وعده‌های غذایی (دقیقه)	9.28	1.91	14.31	6.02	Intake of particles retained on 1.18 mm sieve (kg/d)	مصرف ذرات باقیمانده بر روی الک ۱/۱۸ میلی‌متر (کیلوگرم در روز)
0.092	0.014	0.124	0.069	Eating rate (kg of DM/min)	نرخ خوردن خوراک (کیلوگرم ماده خشک در دقیقه)	5.59	0.85	7.84	4.21	Intake of particles retained on Pan (kg/d)	مصرف ذرات باقیمانده بر روی سینی (کیلوگرم در روز)
1.38	0.40	2.63	0.84	Meal size (kg of DM)	اندازه وعده غذایی (کیلوگرم ماده خشک)	41.71	5.28	52.5	32.9	Milk yield (kg/d)	تولید شیر (کیلوگرم در روز)

نتایج و بحث

مؤثرترین عوامل تأثیرگذار بر درصد چربی شیر که با روش‌های مختلف الگوریتم انتخاب ویژگی به‌دست آمده به تفکیک روش و مدل در جدول ۲ آورده شده است.

جدول ۲- مؤثرترین عوامل تأثیرگذار بر درصد چربی شیر به‌دست آمده با روش‌های مختلف انتخاب ویژگی.

Table 2. The most important factors affecting milk fat percentage obtained by different feature selection methods.

روش Method	مدل Model	فراسنجه Parameter
Best-First	CFS-Subset-Eval	اسیدهای چرب غیر استریفیه خون، کل اسیدهای چرب فرار، عصاره اتری مصرفی، کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم NDF علوفه‌ای مصرفی، کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، وزن بدن، نمره وضعیت بدنی، ضخامت لایه چربی پشت و گلوکز خون
Greedy-Stepwise	CFS-Subset-Eval	اسیدهای چرب غیر استریفیه خون، کل اسیدهای چرب فرار، عصاره اتری مصرفی، کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم NDF علوفه‌ای مصرفی، کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، وزن بدن، نمره وضعیت بدنی، ضخامت لایه چربی پشت و گلوکز خون
Ranker	Info-Gain-Attribute-Eval	اسیدهای چرب غیر استریفیه خون، وزن بدن، نمره وضعیت بدنی، کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، عصاره اتری مصرفی، گلوکز خون، کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، نسبت استات به پروپیونات، کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم NDF علوفه‌ای مصرفی و زمان صرف شده برای استراحت
	Gain-Ratio-Attribute-Eval	اسیدهای چرب غیر استریفیه خون، وزن بدن، نمره وضعیت بدنی، کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، عصاره اتری مصرفی، گلوکز خون، کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، کلسترول خون، ضخامت لایه چربی پشت و مصرف NDF مؤثر فیزیکی برای ذرات درشت‌تر از ۸ میلی‌متر
	Symmetricer-Attribute-Eval	اسیدهای چرب غیر استریفیه خون، وزن بدن، نمره وضعیت بدنی، کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، عصاره اتری مصرفی، گلوکز خون و کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی
	ReliefF-Attribute-Eval	اسیدهای چرب غیر استریفیه خون، وزن بدن، نمره وضعیت بدنی، گلوکز خون، کل زمان صرف شده برای فعالیت نشخواری به ازای هر کیلوگرم NDF علوفه‌ای مصرفی و مصرف ذرات باقیمانده بر روی الک ۱۹ میلی‌متر
	Principal-Components	کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم ماده خشک مصرفی، کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی علوفه‌ای مصرفی، کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم NDF مؤثر فیزیکی مصرفی برای ذرات درشت‌تر از ۱/۱۸ میلی‌متر

در ادامه به‌منظور تولید درخت تصمیم با یک طبقه‌بندی ساده جهت به‌دست آوردن مناسب‌ترین روش انتخاب عوامل مؤثر بر درصد چربی شیر با تعیین آماره‌های ضریب همبستگی، میانگین نسبی خطای مطلق، ریشه میانگین مربع خطا، خطای مطلق نسبی، ریشه مربع خطای نسبی، تعداد طبقات پیش‌بینی شده درست و غلط از الگوریتم‌های Naive Bayes (جدول ۳)، J48 (جدول ۴) و LMT (جدول ۵) استفاده شد که نتایج آن به تفکیک الگوریتم به‌کار رفته و با در نظر گرفتن حالت آزمون 10 Fold و مدل طبقه‌بندی مجموعه آموزش کل (Full training set) در جداول ۳ تا ۵ نمایش داده شده است. در طبقه‌بندی با استفاده از الگوریتم Naive Bayes و J48 روش‌های Best-First و Greedy-Stepwise با بیش‌ترین ضریب همبستگی (۱۰۰ درصد) و بدون تعداد طبقات پیش‌بینی شده غلط به‌عنوان مناسب‌ترین روش‌ها برای تعیین مؤثرترین عوامل تأثیرگذار بر درصد چربی شیر انتخاب شدند.

جدول ۳- ارزیابی طبقات با استفاده از الگوریتم Naive Bayes.

Table 3. The assessment of classes using Naive Bayes algorithm.

روش Method	مدل Model	مدل طبقه‌بندی Classification model	حالت آزمون Test mode	ضریب همبستگی R	میانگین نسبی خطای مطلق ARAE	ریشه میانگین مربع خطاRMSE	خطای مطلق نسبی RAE	ریشه مربع خطای نسبی RSAE	تعداد طبقات پیش‌بینی شده: درست True	تعداد طبقات پیش‌بینی شده: غلط False
Best-First	CFS-Subset-Eval	Full training set	10 Fold class validates	100	0.01	0.06	0.47	3.48	44	0
Greedy-Stepwise	CFS-Subset-Eval	Full training set	10 Fold class validates	100	0.01	0.06	0.49	3.48	44	0
Ranker	Info-Gain-Attribute-Eval	Full training set	10 Fold class validates	97.73	0.20	0.02	5.08	21.33	43	1
	Gain-Ratio-Attribute-Eval	Full training set	10 Fold class validates	95.45	0.20	0.98	5.44	23.37	42	2
	Symmetricer-Attribute-Eval	Full training set	10 Fold class validates	95.45	0.29	0.14	8.19	34.63	42	2
	RelifeF-Attribute-Eval	Full training set	10 Fold class validates	84.09	0.14	0.28	39.38	67.77	37	7
	Principal-Components	Full training set	10 Fold class validates	54.54	0.34	0.48	96.16	115.10	24	20

جدول ۴- ارزیابی طبقات با استفاده از الگوریتم J48.

Table 4. The assessment of classes using J48 algorithm.

روش Method	مدل Model	مدل طبقه‌بندی Classification model	حالت آزمون Test mode	ضریب همبستگی R	میانگین نسبی خطای مطلق ARAE	ریشه میانگین مربع خطا RMSE	خطای مطلق نسبی RAE	ریشه مربع خطای نسبی RSAE	تعداد طبقات پیش‌بینی شده: درست True	تعداد طبقات پیش‌بینی شده: غلط False
Best-First	CFS-Subset-Eval	Full training set	10 Fold class validates	100	0.62	0.23	17.59	54.79	44	0
Greedy-Stepwise	CFS-Subset-Eval	Full training set	10 Fold class validates	100	0.62	0.23	17.59	54.79	44	0
Ranker	Info-Gain-Attribute-Eval	Full training set	10 Fold class validates	90.90	0.817	0.062	0.23	17.59	40	4
	Gain-Ratio-Attribute-Eval	Full training set	10 Fold class validates	90.90	0.62	0.23	17.59	54.79	40	4
	Symmetricer-Attribute-Eval	Full training set	10 Fold class validates	90.90	0.62	0.23	17.59	54.79	40	4
	RelifeF-Attribute-Eval	Full training set	10 Fold class validates	86.36	0.094	0.28	26.38	67.81	38	6
	Principal-Components	Full training set	10 Fold class validates	70.45	0.168	0.35	47.20	85.46	31	13

در طبقه‌بندی با استفاده از الگوریتم LMT، روش‌های Best-First و Greedy-Stepwise با مدل CFS-Subset-Eval و روش Ranker با مدل‌های Gain-Ratio-Attribute-Eval و Symmetricer-Attribute-Eval با بیش‌ترین ضریب همبستگی (۹۷ درصد) و با یک طبقه پیش‌بینی شده غلط به عنوان مناسب‌ترین روش‌ها برای تعیین مؤثرترین عوامل تأثیرگذار بر درصد چربی شیر انتخاب شدند.

در نهایت به‌منظور اطمینان از انتخاب مناسب‌ترین روش برای تعیین مؤثرترین عوامل تأثیرگذار بر درصد چربی شیر از دو روش AMS و ARAE برای تعیین میزان خطای هر یک از روش‌ها استفاده شد که نتایج آن در جدول ۶ نمایش داده شده است. بر این اساس، الگوریتم Naive Bayes با کم‌ترین میزان خطا به‌عنوان مناسب‌ترین الگوریتم برای تعیین عوامل تأثیرگذار بر درصد چربی شیر

انتخاب شد. از این‌رو، داده مربوط به اسیدهای چرب غیر استریفیه خون، کل اسیدهای چرب فرار، عصاره اتری مصرفی، کل زمان صرف شده برای فعالیت نشخوار به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، کل زمان صرف شده برای فعالیت نشخوار به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی علوفه‌ای مصرفی، کل زمان صرف شده برای فعالیت جویدن به ازای هر کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی، وزن بدن، نمره وضعیت بدنی، ضخامت لایه چربی پشت و گلوکز خون به‌عنوان بهترین و مؤثرترین عوامل تأثیرگذار بر چربی شیر از بین ۶۶ گروه داده ورودی به مدل انتخاب شدند. منبع اسیدهای چرب از پیش‌ساخته شده شیر (اسیدهای چرب بلند زنجیر با بیش از ۱۸ اتم کربن) اسیدهای چربی هستند که از طریق خوراک (عصاره اتری مصرفی) و بافت چربی (با از دست دادن وزن بدن و کم کردن نمره وضعیت بدنی و ضخامت لایه چربی پشت) به‌صورت تری‌گلیسرید و اسیدهای چرب غیر استریفیه وارد خون شده و از آن طریق به بافت پستانی راه پیدا می‌کنند (۴ و ۱۲).

جدول ۵- ارزیابی طبقات با استفاده از الگوریتم LMT.

Table 5. The assessment of classes using LMT algorithm.

روش Method	مدل Model	مدل طبقه‌بندی Classification model	حالت آزمون Test mode	ضریب همبستگی R	میانگین نسبی خطای مطلق ARAE	ریشه میانگین مربع خطا RMSE	خطای مطلق نسبی RAE	ریشه مربع خطای نسبی RSAE	تعداد طبقات پیش‌بینی شده: درست True	تعداد طبقات پیش‌بینی شده: غلط False
Best-First	CFS-Subset-Eval	Full training set	10 Fold class validates	97.72	0.66	0.14	18.69	33.85	43	1
Greedy-Stepwise	CFS-Subset-Eval	Full training set	10 Fold class validates	97.72	0.66	0.14	18.69	33.85	43	1
Ranker	Info-Gain-Attribute-Eval	Full training set	10 Fold class validates	93.18	0.10	0.20	26.90	48.48	41	3
	Gain-Ratio-Attribute-Eval	Full training set	10 Fold class validates	97.71	0.24	0.10	6.95	23.89	43	1
	Symmetricer-Attribute-Eval	Full training set	10 Fold class validates	97.70	0.03	0.12	9.68	28.75	43	1
	RelifeF-Attribute-Eval	Full training set	10 Fold class validates	84.09	0.20	0.31	55.21	73.81	37	7
	Principal-Components	Full training set	10 Fold class validates	84.09	0.18	0.31	50.58	74.82	37	7

جدول ۶- دقت طبقه‌های مختلف الگوریتم انتخاب ویژگی.

Table 6. The accuracy of different classes of Feature selection algorithm.

روش ارزیابی Evaluation method						مدل Model	روش Method
Naive Bayes		J48		LMT			
AMS	ARAE	AMS	ARAE	AMS	ARAE		
0.00	1.11	0.00	39.98	1.54	42.48	CFS-Subset-Eval	Best-First
0.00	1.11	0.00	39.98	1.54	42.48	CFS-Subset-Eval	Greedy-Stepwise
1.54	11.55	6.15	0.52	1.54	61.14	Info-Gain-Attribute-Eval	Ranker
3.08	12.36	6.15	39.98	1.54	15.80	Gain-Ratio-Attribute-Eval	
3.08	18.61	6.15	39.98	1.54	22.00	Symmetricer-Attribute-Eval	
10.77	89.50	9.23	59.95	1.54	125.48	RelifeF-Attribute-Eval	
30.77	218.55	20.00	107.27	1.54	114.95	Principal-Components	

بخشی از گلوکزی که وارد بافت پستانی می‌شود برای ساخت لاکتوز مورد استفاده قرار می‌گیرد و مابقی آن می‌تواند در جهت تولید انرژی برای افزایش ساخت NADPH از مسیر پنتوز فسفات یا از تبدیل ایزوسیترات به ۲- اُگزوگلوتارات به‌کار رود. از این‌رو، مولکول‌های NADPH ساخته شده از این مسیرها می‌تواند در ساخت چربی شیر در بافت پستانی مورد استفاده قرار بگیرد (۱۲). بر اساس نتایج به‌دست آمده توجه به فعالیت نشخوار گاوها بر مبنای کیلوگرم الیاف نامحلول در شوینده خنثی مصرفی و الیاف نامحلول در شوینده خنثی علوفه‌ای مصرفی حائز اهمیت است و بایستی که از مقدار بهینه الیاف نامحلول در شوینده خنثی و الیاف نامحلول در شوینده خنثی علوفه‌ای در جیره گاوهای شیری اطمینان حاصل کرد تا از فعالیت نشخواری گاوها کاسته نشود. توجه به این نکته ضروری است که مدیریت ضعیف و نامناسب محیط پرورش می‌تواند فعالیت نشخوار گاوها را حتی با حفظ مقدار بهینه الیاف در جیره کاهش بدهد. فعالیت نشخوار طی یک دوره ۲۴ ساعته صورت می‌گیرد و معمولاً تحت شرایط ایده‌ال گاوهای بالغ ۴۸۰ تا ۵۴۰ دقیقه را در شبانه روز صرف آن می‌کنند (۱۳). طیف وسیعی از عوامل مدیریتی مانند تراکم زیاد (۱۰ تا ۲۰ درصد)، بهاربندهایی با آمیزه‌ای از گاوهای تک و

چند شکم‌زا (۱۵ درصد)، زیادی زمان صرف شده در گردن‌گیرها (۱۴ درصد) و تنش گرمایی (۱۰ تا ۲۲ درصد) می‌توانند فعالیت نشخوار را کاهش بدهند. اگر فعالیت نشخوار به طور مزمن ۱۰ تا ۲۰ درصد به سبب مدیریت ضعیف کاهش پیدا بکند بر آن اساس می‌توان به مخاطره افتادن عملکرد شکمبه و افزایش خطر مشکلاتی مانند اسیدوز شکمبه‌ای تحت حاد، بازده گوارشی ضعیف‌تر، لنگش و در نهایت کاهش خروج چربی و پروتئین از شیر را انتظار داشت (۲). این کاهش در چربی شیر می‌تواند با نامتوازنی در تولید اسیدهای چرب فرار و یا تولید واسطه‌های زیست‌هیدروژنه‌ای کاهش دهنده چربی شیر در ارتباط باشد (۴، ۷، ۸ و ۹).

نتیجه‌گیری

بر اساس انتخاب تعداد محدودی از عوامل مؤثر بر چربی شیر با الگوریتم انتخاب ویژگی امکان پایش و اندازه‌گیری آن‌ها جهت حفظ مؤثر چربی شیر در مزرعه و صرفه‌جویی در زمان و هزینه وجود دارد.

منابع

1.Dash, M., and Liu, H. 2003. Consistency-based search in feature selection. Artif. Intel. 151: 155–176.

2.Grant, R.J., and Dann, H.M. 2015. Biological importance of rumination and its use on-farm, Cornell Nutrition Conference for Feed Manufacturers, Cornell, USA, Pp: 1–10.

3.Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H. 2015. The WEKA Data Mining Software Version 3.8. The University of Waikato, Hamilton, New Zealand.

4.Kargar, S., Ghorbani, G.R., Alikhani, M., Khorvash, M., Rashidi, L., and Schingoethe, D.J. 2012. Lactational performance and milk fatty acid profile of Holstein cows in response to dietary fat supplements and forage:concentrate ratio. J. Livest. Sci. 150: 274–283.

5.Kargar, S., Ghorbani, G.R., Fievez, V., and Schingoethe, D.J. 2015a. Performance, bioenergetic status, and indicators of oxidative stress of environmentally heat-loaded Holstein cows in response to diets inducing milk fat depression. J. Dairy Sci. 98: 4772–4784.

6.Kargar, S., Ghorbani, G.R., and Khorvash, M. 2015b. Nutrient digestibility and rumen fermentation parameters in response to dietary grain source and oil supplement of Holstein dairy cows. Ir. J. Anim. Sci. Res. 7: 285–293. (In Persian)

7.Kargar, S., Ghorbani, G.R., Khorvash, M., Kahyani, A., Karimi-Dehkordi, S., Safahani-Langarudi, M., Fievez, V., and Schingoethe, D.J. 2017. Milk fat secretion in Holstein dairy cows: Insights from grain type and oil supplement.J. Livest. Sci. 196: 36–41.

8.Kargar, S., Ghorbani, G.R., Khorvash, M., Kamalian, E., and Schingoethe, D.J. 2013. Dietary grain source and oil supplement: Feeding behavior and lactational performance of Holstein cows.J. Livest. Sci. 157: 162–172.
9.Kargar, S., Ghorbani, G.R., Khorvash, M., Sadeghi-Sefidmazgi, A., and Schingoethe, D.J. 2014. Reciprocal combinations of barley and corn grains in oil-supplemented diets: Feeding behavior and milk yield of lactating cows J. Dairy Sci. 97: 7001–7011.
10.Landwehr, N., Hall, M., and Frank, E. 2005. Logistic model trees. Mach. Learn. 59: 161–205.
11.Naseriparsa, M., Bidgoli, A.M., and Varaee, T. 2014. A hybrid feature selection method to improve performance of a group of classification algorithms. Int. J. Comput. Appl. 69: 28–35.
12.Rigout, S., Hurtaud, C., Lemosquet, S., Bach, A., and Rulquin, H. 2003. Lactational effects of propionic acid and duodenal glucose in cows. J. Dairy Sci. 86: 243–253.
13.Van Soest, P.J. 1994. Nutritional ecology of the ruminant. Cornell University Press, Ithaca, NY, USA.

Gorgan University of Agricultural Sciences and Natural Resources
J. of Ruminant Research, Vol. 4(4), 2016
http://ejrr.gau.ac.ir

Use of feature selection algorithm to determine the most important factors affecting milk fat percentage of Holstein cows

***S. Kargar[1] and M. Mokarram[2]**
[1]Assistant Prof., Dept. of Animal Sciences, faculty of Agriculture, Shiraz University,
[2]Assistant Prof., Dept. of Range and Watershed Management, Faculty of Agriculture and Natural Resources of Darab, Shiraz University, Darab, Shiraz, Iran
Received: 11/03/2016; Accepted: 01/18/2017

Abstract

Background and objectives: Monitoring the milk components of a herd can help assess the health and nutritional status of lactating cows. Fat is one of the most valuable components of milk. Milk fat (proportion and total amount) is typically the most variable component in milk which is influenced by various physiological and environmental factors. A herd milk fat test below 0.3% of breed average can indicate a problem situation. Therefore, diets that allow for an improvement in milk fat output would potentially be economically advantageous. In order to maintain milk fat, it is necessary to identify, monitor, and measure the most important factors influencing milk fat. Feature selection algorithm is one of data mining methods and to the best of our knowledge no available literature to use this algorithm to refine most contributing factors affecting milk fat in Holstein dairy cows. Objective of this study was to determine the most important factors affecting milk fat of Holstein cows using feature selection algorithm.

Materials and methods: In the current study, a total of 2112 raw data obtained from 66 factors (which might potentially able to affect milk fat) was used. Data was evaluated by the three important methods (Best-First, Greedy-Stepwise, and Ranker methods) and six models (CFS-Subset-Eval, Info-Gain-Attribute-Eval, Gain-Ratio-Attribute-Eval, Symmetricer-Attribute-Eval, RelifeF-Attribute-Eval, Principal-Components) of feature selection algorithm. Data was analyzed using the WEKA Software (v. 3.8).

Results: Results indicated that both Best-First and Gready-Stepwise methods of feature selection algorithm are the most appropriate methods to select efficient

*Correspondig author: skargar@shirazu.ac.ir

factors affecting milk fat using Naive Bayes classification with minimized error. Accordingly, blood non-esterified fatty acid, total volatile fatty acids of rumen fluid, ether-extract intake, time spent rumination per kilogram of neutral detergent fiber intake, time spent rumination per kilogram of forage neutral detergent fiber intake, total chewing time per kilogram of neutral detergent fiber intake, body weight, body condition score, back fat thickness, and blood glucose were the most important factors affecting milk fat.

Conclusion: It is possible to monitor and measure of some limited factors selected by feature selection to efficiently maintain milk fat on-farm and save time and cost.

Keywords: Milk fat, Feature selection algorithm, Dairy cow