چکیده

در سالیان گذشته بعدازاینکه چالش‌های مربوط به جمع‌آوری داده به نحوی مرتفع گشتند، اکنون سؤال اصلی به چگونگی پردازش بر روی این حجم عظیم از داده‌ها تبدیل شده است. برای بهره‌برداری و اکتشاف دانش از این حجم بسیار زیاد داده، ضروری است که ابزارها و زیرساخت‌های مناسب آن ایجاد و استفاده گردد. طبق مدل ۵ مرحله‌ای که برای فرآیند استخراج و اکتشاف دانش مورد استفاده قرار می‌گیرد، مهم‌ترین و اساسی‌ترین قسمت، پیش‌پردازش و آماده‌سازی داده است که نتیجه آن کیفیت خروجی و مراحل بعدی مدل را نیز ضمانت می‌کند. زمانی که منابع داده زیادی اعم از برخط (جریانی) و برون‌خط (غیرجریانی) وجود داشته باشند، چگونگی پیش‌پردازش داده‌ها، پاک‌سازی داده‌ها، ادغام داده‌ها، تبدیل داده‌ها و ذخیره-سازی قالب‌های مشخص از داده‌ها مسائلی هستند که با آن روبرو هستیم. هدف از چارچوب نظام‌مند، ایجاد ساختاری برای تعریف انواع منابع ورودی، انواع نیازمندی‌های داده‌ای، انواع مؤلفه‌های پیش‌پردازش، انواع شروط داده‌ای و از این قبیل عملکردها می‌باشد. درواقع از طریق این چارچوب کاربر می‌تواند ساختارهای مختلف و پیچیده‌ای را برای نیازهای خاص خود به‌منظور تعیین ساختار پیش‌پردازش برای منابع داده‌ای مختلف جریانی و غیرجریانی ایجاد و تعریف کند. در این رساله یک چارچوب نظام‌مند برای پیش‌پردازش داده‌های جریانی و غیرجریانی، یک الگوریتم هوشمند برای اجرای روال پیش‌پردازش تعریف‌شده توسط کاربر با توجه به منابع در اختیار و شروط و نیازمندی‌های تعریف‌شده از سوی کاربر و همچنین یک الگوریتم تطبیق برای ایجاد امکان خوداصلاحی و تطبیق اجرای روال پیش‌پردازش کاربر ارائه شده است. برای ارزیابی چارچوب پیشنهادی از مجموعه‌داده حوزه شهر هوشمند و اینترنت اشیا، استفاده شده است. چارچوب پیشنهادی از دو منظر برخط و غیربرخط مورد ارزیابی قرار گرفت و در مواردی که امکان مقایسه با سایر کارها وجود داشته باشد نیز با روش‌های به‌روز مقایسه انجام گردید. روش پیشنهادی در قالب سناریوهای مختلف مورد ارزیابی قرار گرفت. به‌عبارت‌دیگر با استفاده از چارچوب پیشنهادی کاربردهای مختلف در حوزه داده‌های حجیم پیاده‌سازی شدند تا تأثیر استفاده از چارچوب پیشنهادی و مؤلفه زمان‌بند در کاهش زمان اجرا و کم کردن پیچیدگی پیاده‌سازی مشخص گردد. نتایج آزمایشات در سه سناریوی مختلف نشان دهنده عملکرد مناسب چارچوب پیشنهادی از نظر کاربرد و تسریع در زمان اجرای کارها می‌باشد.