

با افزایش حجم داده‌های تولید شده و عدم توانایی پاسخگویی بهنگام سرویس‌دهنده‌ها، محاسبات تقریبی بخش اجتناب ناپذیری از تحلیل‌های کلان‌داده شده است. در این پژوهش قصد داریم تا با استفاده از روش‌های آماری و به کمک نمونه‌برداری، خاصه نمونه‌برداری اهمیت، علاوه بر تقریب نتیجه‌ی مورد نظر، واریانس این تقریب نیز با توجه به محدودیت‌های مساله به حداقل مقدار ممکن برسد. استفاده از نمونه‌برداری در تحلیل داده، بدون در نظر گرفتن توزیع آماری آن ممکن است نتیجه‌گیری را منحرف و بعضا کاملا اشتباه نماید. چنین شرایطی زمانی تاثیرگذارتر خواهد بود که داده‌های مهم کمیاب بوده و دستیابی به آن‌ها نیازمند بررسی تمام داده‌های موجود باشد. در این پایان‌نامه با بررسی توزیع مجموعه‌ی داده و تخمین پارامترهای آن، توزیع جایگزینی بدست می‌آوریم که با استفاده از آن احتمال انتخاب عناصر مهم این مجموعه افزایش خواهد یافت، که استفاده از آن‌ها می‌تواند نتیجه‌ی نهایی را بهبود دهد. در نهایت با استفاده از روش‌های آماری میزان خطای پارامتر محاسبه شده را بصورت یک بازه‌ی اطمینان مطرح خواهیم نمود. نتایج بدست آمده نشان می‌دهد که با استفاده از این روش زمان مورد نیاز برای نمایش نتیجه به کاربر تا ۶۶٪ برای کاربردهای ساده و تا ۴۰٪ برای کاربردهای تکرارشونده، قابل کاهش است.