با رشد و گسترش رایانه های شخصی و گوشی های تلفن همراه هوشمند، توسعه شبکه های حسگر و اینترنت اشیا، فراگیر شدن اشتراک گذاری مطالب توسط کاربران و توسعه محاسبات ابری حجم عظیمی از داده ها توسط افراد ایجاد می شود. نحوه مدیریت این داده ها که حجم، سرعت تولید و تنوع زیادی دارند سبب شده است که موضوع داده های عظیم به یکی از چالش برانگیزترین موضوعات روز بدل شود. در کنار این رشد سریع تولید داده، راه حل هایی برای مدیریت آن معرفی شده است. هدوپ به عنوان یکی از راه حل های محبوب برای کار با داده های عظیم، چارچوبی را برای توزیع داده ها و مدیریت منابع توزیع شده فراهم می آورد.هدوپ با استفاده از امکانات ذکر شده، چارچوب نگاشت-کاهش را برای کاربر مهیا می سازد. در روش های یادگیری ماشینی، به طور ویژه یادگیری عمیق، استفاده از داده های بیشتر به کیفیت بهتر ماشین کمک خواهد کرد. یکی از چالش های پیش رو در این حوزه، استفاده از دادگان حجیم موجود بدون افزایش بیش از حد زمان مورد نیاز برای آموزش و آزمودن مدل است. با این حال الگوریتم های ارائه شده اندکی تا بحال معرفی شده اند و غالب کارهای موجود به کمک روش های تقریبی به حل این مسئله پرداخته اند. هرچند چنین رویکردی از باب این که در یادگیری ماشینی کیفیت نهایی کار اهمیت دارد، مورد قبول است لیکن چنین دیدگاهی تنها به تسریع عمل یادگیری ماشین می پردازد و از بهبود کیفیت آن غافل است. به طور مثال در شبکه حافظه ای [۱] ماشین نیازمند جستجو در حافظه ای عظیم از بردارهاست. در چنین شرایطی با رشد حجم داده این شبکه کارایی خود را عملا از دست خواهد داد. به عنوان مثالی دیگر، راه حل هایی برای یادگیری شبکه های عصبی عمیق به کمک پردازش موازی معرفی شده است [۲]، اما این روش ها از دارای هزینه محاسباتی و پیچیدگی مدل بالا می باشند. در این پژوهش قصد داریم که با استفاده از محاسبات توزیع شده بر پایه چارچوب نگاشت-کاهش که در هدوپ مهیا شده است برخی از روش های یادگیری ماشینی غیرعمیق و عمیق را به صورت توزیع شده ارائه دهیم و به بررسی میزان بهبود حاصل از روش های ارائه شده بپردازیم. به طور خاص دو روش نزدیک ترین همسایه ها وشبکه های عصبی تنیده شده به عنوان دو مثال از روش های یادگیری به ترتیب غیرعمیق و عمیق انتخاب شده اند و برای هر کدام به همراه روش توزیع شده، آزمایشات متعددی انجام شده است. نتایج حاصل از آزمایشات انجام شده در این پژوهش نشان می دهند که روش های ارائه شده به خوبی قادر خواهند بود بر روی دادگان با حجم بالا کار کنند.